



República Federativa do Brasil
Ministério da Indústria, Comércio Exterior
e Serviços
Instituto Nacional da Propriedade Industrial

(11) PI 0304565-0 B1

(22) Data do Depósito: 14/03/2003

(45) Data de Concessão: 24/07/2018



(54) Título: MÉTODO DE PREDIÇÃO DE MOVIMENTO COMPENSADO PARA USO NA CODIFICAÇÃO DA SEQUÊNCIA DE VÍDEO DIGITAL, CODIFICADOR E DECODIFICADOR DE VÍDEO PARA CODIFICAR/DECODIFICAR UMA SEQUÊNCIA DE VÍDEO DIGITAL USANDO PREDIÇÃO DE MOVIMENTO COMPENSADO, TERMINAL DE MULTIMÍDIA, E, CODEC DE VÍDEO

(51) Int.Cl.: H04N 5/91

(30) Prioridade Unionista: 15/03/2002 US 60/365,072

(73) Titular(es): NOKIA TECHNOLOGIES OY

(72) Inventor(es): JANI LAINEMA

(85) Data do Início da Fase Nacional: 26/12/2003

“MÉTODO DE PREDIÇÃO DE MOVIMENTO COMPENSADO PARA USO NA CODIFICAÇÃO DA SEQUÊNCIA DE VÍDEO DIGITAL, CODIFICADOR E DECODIFICADOR DE VÍDEO PARA CODIFICAR/DECODIFICAR UMA SEQUÊNCIA DE VÍDEO DIGITAL USANDO PREDIÇÃO DE MOVIMENTO COMPENSADO, TERMINAL DE MULTIMÍDIA, E, CODEC DE VÍDEO”.

Campo da Invenção

A invenção se refere em termos gerais aos sistemas de comunicação e mais especificamente a compensação de movimento na codificação de vídeo.

Descrição da Técnica Anterior

Uma sequência de vídeo digital, tal como uma imagem em movimento ordinário gravada em um filme comum, compreende uma sequência de imagens imóveis, sendo a ilusão de movimento criada pela apresentação sucessiva de imagens consecutivas da sequência a uma velocidade relativamente alta, tipicamente de 15 a 30 quadros por segundo.

Devido à velocidade de apresentação relativamente alta dos quadros, as imagens nos quadros consecutivos têm a tendência de serem bem semelhantes e contêm, portanto, uma quantidade considerável de informações redundantes. Uma cena típica pode compreender, por exemplo, alguns elementos estacionários, tais como cenário de fundo, e algumas áreas móveis, que podem assumir muitas formas diferentes, o rosto de um leitor de jornal, por exemplo, tráfego em movimento e assim por diante. Alternativamente, ou como um acréscimo, pode se encontrar um movimento conhecido como "movimento global" na sequência de vídeo, devido, por exemplo, a translação, a efeito panorâmico ou um zoom da câmara que registra a cena. No entanto, em muitos casos a alteração total de um quadro de vídeo para o outro é bastante pequena.

Cada quadro de uma sequência de vídeo digital não compactada compreende um conjunto de pixels de imagens. Em um formato de vídeo digital habitualmente empregado conhecido como Formato de Intertroca Comum Quaternário (QCIF – Quarter Common Interchange Format), um quadro compreende um conjunto de 176 x 144 pixels, tendo neste caso cada quadro 25.344 pixels. Por sua vez, cada pixel é representado por um determinado número de bits que transportam informações sobre o conteúdo de luminância



e/ou de cor da região da imagem que corresponde ao pixel. Habitualmente é usado um modelo de cor denominado YUV para representar o conteúdo de luminância/crominância da imagem. A luminância ou componente Y, representa a intensidade (brilho) da imagem, ao passo que o conteúdo de cor da imagem é representado por dois componentes de crominância ou de diferença entre cores, denominados de U e V.

Os modelos de cores baseados em uma representação de luminância/crominância do conteúdo da imagem proporcionam certas vantagens quando comparados com os modelos de cores que são baseados em uma representação que abrange cores primárias (isto é Vermelho, Verde e Azul, RGB). O sistema visual humano é mais sensível às variações de intensidade do que às variações de cor e, os modelos de cores YUV exploram esta propriedade usando uma resolução espacial menor para os componentes de crominância (U, V) do que para o componente de luminância (Y). Deste modo pode ser reduzido o número de informações necessárias para se codificar as informações de cores em uma imagem, e com uma redução aceitável na qualidade de imagem.

A resolução espacial inferior dos componentes de crominância é geralmente obtida pela sub-amostragem espacial. Tipicamente, cada quadro de uma sequência de vídeo é dividido em unidades denominadas "macroblocos" que compreendem informações de luminância (Y) e informações de crominância (U, V) associadas (especialmente sub-amostradas). A Figura 3 ilustra um modo no qual os macroblocos podem ser formados. A Figura 3a mostra um quadro de uma sequência de vídeo representado usando-se um modelo de cores YUV, tendo cada componente a mesma resolução espacial. Os macroblocos são formados pela representação de uma região de 16 x 16 pixels de imagem na imagem original (Figura 3b) em forma de quatro blocos de informações de luminância, compreendendo cada bloco de luminância um conjunto de 8 x 8 valores de luminância (Y) e dois componentes de crominância espacialmente correspondentes (U e V) que são sub-amostrados por um fator de dois nas direções horizontal e vertical, resultando nos conjuntos correspondentes de 8 x 8 valores de crominância (U,V) (veja Figura 3c).

Uma imagem QCIF compreende 11 x 9 macroblocos. Se os blocos de luminância e os blocos de crominância são representados com uma resolução de 8 bits (isto é, por números na faixa de 0 a 255), o número total de bits necessário por macrobloco é (16

$x 16 \times 8) + 2 \times (8 \times 8 \times 8) = 3072$ bits. O número de bits necessário para representar um quadro de vídeo no formato QCIF é, portanto de $99 \times 3072 = 304.128$ bits. Isto significa que a quantidade de dados necessários para transmitir/registrar/apresentar uma seqüência de vídeo não compactada no formato QCIF, representada empregando-se um modelo de cores

YUV, a uma velocidade de 30 quadros por segundo é de mais de 9 Mbps (milhões de bits por segundo). Isto é uma velocidade de dados extremamente alta e impraticável para ser empregada em aplicações de registro, transmissão e apresentação de vídeo devido às exigências demasiado grandes de capacidade de armazenagem e capacidade de canais de transmissão e de desempenho dos componentes físicos. Se os dados de vídeo tiverem que ser transmitidos em tempo real através de uma rede de linhas fixas tais como a RDSI (Rede Digital de Serviços Integrados) ou uma RTCP convencional (Rede de Telefonia Comutada Pública), a largura de banda de transmissão de dados disponível é tipicamente da ordem de 64 kbits/s. Em vídeo telefonia móvel, onde a transmissão tem lugar pelo menos em parte através de uma conexão de comunicação por rádio, a largura de banda disponível pode ser menor ainda, de até 20 kbits/s. Isso significa que deve ser obtida uma redução significativa na quantidade de informações empregada para representar os dados de vídeo para permitir uma transmissão de seqüências de vídeo através de redes de comunicação de larguras de banda estreitas. Por este motivo foram desenvolvidas técnicas de compressão de vídeo que reduzem a quantidade de informações transmitida e que conservam ao mesmo tempo uma qualidade de imagem aceitável.

Os processos de compressão de vídeo são baseados na redução das partes redundantes e irrelevantes do ponto de vista da percepção das seqüências de vídeo. A redundância nas seqüências de vídeo pode ser distribuída em categorias de redundância espacial, temporal e espectral. "Redundância espacial" é o termo usado para descrever a correlação (similaridade) entre pixels vizinhos no interior do quadro. O termo "redundância temporal" expressa o fato de que objetos que aparecem em um quadro de uma seqüência têm a probabilidade de aparecer em quadros subseqüentes, ao passo que "redundância espectral" se refere à correlação entre componentes diferentes de cor da mesma imagem.

Uma compressão suficientemente eficiente não pode ser habitualmente obtida por uma simples redução das diversas formas de redundância em uma seqüência dada



de imagens. Assim, os codificadores de vídeo mais correntes também reduzem a qualidade daquelas partes da seqüência que são subjetivamente menos importantes. Além disso, a redundância da corrente de bits de vídeo compactada em si é reduzida por meio de uma codificação sem-perdas eficiente. Geralmente, isto se consegue empregando-se uma técnica conhecida como codificação com entropia.

Existe freqüentemente uma quantidade significativa de redundância espacial entre os pixels que compõe cada quadro de uma seqüência de vídeo digital. Em outras palavras, o valor de qualquer pixel no interior de um quadro da seqüência é substancialmente o mesmo que o valor de outros pixels na sua proximidade imediata.

Tipicamente, os sistemas de codificação de vídeo reduzem a redundância espacial empregando uma técnica conhecida como "codificação de transformação com base em blocos" em que uma transformação matemática, tal como a Transformada de Coseno Discreto Bidimensional (DCT) é aplicada a blocos de pixels de imagem. Isto transforma os dados de imagem de uma representação compreendendo valores em pixels em uma forma que compreende um conjunto de valores de coeficientes representativos dos componentes de freqüência espacial, reduzindo significativamente a redundância espacial e produzindo assim uma representação mais compacta dos dados de imagem.

Os quadros de uma seqüência de vídeo que são comprimidos empregando-se uma codificação de transformação com base em blocos, sem referência a qualquer outro quadro no interior da seqüência, são denominados INTRA-codificados ou quadros I. Além disso, e sempre que possível, os blocos de quadros INTRA-codificados são previstos a partir de blocos anteriormente codificados no interior do mesmo quadro. Esta técnica, conhecida como INTRA-predição tem o efeito de reduzir ainda mais a quantidade de dados necessária para representa um quadro INTRA-codificado.

Geralmente, os sistemas de codificação de vídeo não somente reduzem a redundância especial no interior de quadros individuais de uma seqüência de vídeo, como também utilizam uma técnica conhecida como "predição de movimento compensado", para reduzir a redundância temporal na seqüência. Empregando-se a predição de movimento compensado, o conteúdo da imagem de alguns (freqüentemente de muitos) quadros em uma seqüência de vídeo digital é "previsto" a partir de um ou mais outros quadros na seqüência,



conhecidos como quadros "de referência". A predição do conteúdo da imagem é obtido por rastreamento do movimento de objetos ou regiões de uma imagem entre um quadro a ser codificado (comprimido) e o(s) quadro(s) de referência empregando-se "vetores de movimento". Em geral, o(s) quadro(s) de referência pode(m) preceder o quadro a ser codificado ou pode(m) segui-lo na sequência de vídeo. Como no caso da codificação-INTRA, a predição de movimento compensado de um quadro de vídeo é tipicamente executada macrobloco a macrobloco.

Os quadros de uma sequência de vídeo que são comprimidos empregando-se predição de movimento compensado são geralmente denominados INTER-codificados ou quadros P. A predição de movimento compensado sozinha raramente proporciona uma representação suficientemente precisa do conteúdo da imagem de um quadro de vídeo e portanto é tipicamente necessário se fornecer um quadro denominado quadro de "erro de predição" (PE) com cada quadro INTER-codificado. O quadro com erro de predição representa a diferença entre uma versão decodificada do quadro INTER-codificado e o conteúdo de imagem do quadro a ser codificado. Mais especificamente, o quadro com erro de predição compreende valores que representam a diferença entre os valores em pixels do quadro a ser codificado e os valores em pixels correspondentes reconstruídos formados com base em uma versão prevista do quadro em questão. Conseqüentemente, o quadro com erro de predição tem características análogas a uma imagem imóvel e pode ser aplicada a codificação de transformação com base em blocos para se reduzir sua redundância espacial e, portanto a quantidade de dados (número de bits) necessários para representá-lo.

A fim de se ilustrar a operação de um sistema genérico de codificação de vídeo com mais detalhes, far-se-á referência agora ao codificador de vídeo e ao decodificador de vídeo, exemplos estes ilustrados nas Figuras 1 e 2 dos desenhos apensos. O codificador de vídeo **100** da Figura 1 emprega uma combinação de Codificação-INTRA e Codificação-INTER para produzir uma corrente de bits de vídeo compactados (codificados) e o decodificador **200** da Figura 2 é disposto para receber e decodificar a corrente de bits de vídeo produzida pelo codificador **100** para produzir uma sequência de vídeo reconstruída. Em toda a descrição que segue será pressuposto que o componente de luminância de um macrobloco compreende 16 x 16 pixels dispostos em forma de um conjunto de 4 blocos de 8

x 8, e que os componentes de crominância associados são espacialmente sub-amostrados por um fator de dois na direção horizontal e na vertical para formar blocos de 8 x 8, conforme ilustrado na Figura 3. A extensão da descrição a outros tamanhos de blocos e a outros esquemas de sub-amostragem se tornará aparente aos peritos na técnica.

5 O codificador de vídeo **100** compreende uma entrada **101** para receber um sinal de vídeo digital de uma câmara ou de outra fonte de vídeo (não mostrada). Ele também compreende uma unidade de transformação **104** que é disposta de modo a executar uma transformada de cossenos discretos com base nos blocos (DCT), um quantizador **106**, um quantizador inverso **108**, uma unidade de transformação inversa **110**, disposta para executar uma transformada inversa de cossenos discretos com base em blocos (IDCT), combinadores 10 **112 e 116** e uma memória de quadros **120**. O codificador compreende ainda um avaliador de movimento **130**, um codificador de campo de movimento **140** e um preditor de movimento compensado **150**. Os comutadores **102 e 114** são operados em regime de cooperação pelo gerenciador de controle **160** para comutar o codificador entre o modo INTRA de 15 codificação de vídeo e o modo INTER de codificação de vídeo. O codificador **100** também compreende um codificador multiplexador de vídeo **170** que forma uma única corrente de bits a partir de diversos tipos de informações produzidas pelo codificador **100** para uma transmissão subsequente a um terminal receptor remoto ou, para a armazenagem em um meio de armazenagem em massa, por exemplo, tal como um disco rígido de computador 20 (não mostrado).

O codificador **100** opera do seguinte modo. Cada quadro de vídeo não comprimido fornecido da fonte de vídeo à entrada **101** é recebido e processado macrobloco a macrobloco, de preferência na ordem de varredura de rastreamento. Quando a codificação de uma nova seqüência de vídeo começa, o primeiro quadro a ser codificado é codificado 25 como um quadro INTRA-codificado. Subseqüentemente, o codificador é programado para codificar cada quadro em formato INTER-codificado, a não ser que uma das seguintes condições sejam satisfeitas: 1) considera-se que o macrobloco corrente do quadro que está sendo codificado é tão diferente dos valores em pixels no quadro de referência usado na predição que é produzida uma quantidade excessiva de informações de erro de predição, 30 sendo neste caso o macrobloco corrente codificado no formato INTRA-codificado; 2) um

intervalo de repetição de quadro INTRA predefinido tenha expirado; ou 3) recebe-se um retorno de um terminal de recepção indicando uma solicitação para que um quadro seja fornecido em formato INTRA-codificado.

A ocorrência da condição 1) é detectada por monitoração da saída do combinador **116**. O combinador **116** forma uma diferença entre o macrobloco atual do quadro que está sendo codificado e a sua predição produzida no bloco de predição de movimento compensado **150**. Se uma medida desta diferença (uma soma de diferenças absolutas de valores em pixels, por exemplo) exceder um limiar predeterminado, o combinador **116** informa o gerenciador de controle **160** através de uma linha de controle **119** e o gerenciador de controle **160** opera os comutadores **102** e **114** através da linha de controle **113** para comutar o codificador **100** para o modo de Codificação-INTRA. Deste modo, um quadro que seria caso contrário codificado no formato INTER-codificado pode compreender macroblocos INTRA-codificados. A ocorrência da condição 2) é monitorada por meio de um cronômetro ou contador de quadros implementado no gerenciador de controle **160**, de modo tal que se o cronômetro expirar, ou se o contador de quadros atingir um número predeterminado de quadros, o gerenciador de controle **160** opera os comutadores **102** e **114** através da linha de controle **113** para comutar o codificador para o modo de Codificação-INTRA. A condição 3) é desencadeada se o gerenciador de controle **160** receber um sinal de retorno de um terminal receptor, através da linha de controle **121**, por exemplo, indicando que é necessária uma recapitulação de quadro INTRA pelo terminal de recepção. Tal condição pode ocorrer, por exemplo, se um quadro anteriormente transmitido for muito corrompido por interferência durante a sua transmissão, tornando impossível a sua decodificação no receptor. Nesta situação, o decodificador receptor emite uma solicitação para que o quadro seguinte seja codificado em formato INTRA-codificado, reiniciando, assim, a seqüência de codificação.

Será agora descrita a operação do codificador **100** no modo de Codificação-INTRA. No modo de Codificação-INTRA, o gerenciador de controle **160** opera o comutador **102** para aceitar a entrada de vídeo da linha de entrada **118**. A entrada de sinal de vídeo é recebida macrobloco a macrobloco da entrada **101** através da linha de entrada **118**. À medida que eles são recebidos, os blocos de valores de luminância e croma-
nância que



constituem o macrobloco são passados ao bloco de transformação DCT **104**, que executa uma transformação de cossenos discretos bidimensionais em cada bloco de valores, produzindo um conjunto bidimensional de coeficientes DCT para cada bloco. O bloco de transformação DCT **104** produz um conjunto de valores de coeficientes para cada bloco, correspondendo o número de valores de coeficientes às dimensões dos blocos que compõem o macrobloco (neste caso 8×8). Os coeficientes DCT para cada bloco são passados ao quantizador **106**, onde eles são quantizados utilizando-se o parâmetro de quantização QP. A seleção do parâmetro de quantização QP é controlada pelo gerenciador de controle **160** através da linha de controle **115**.

O conjunto de coeficientes DCT quantizados para cada bloco é então passado do quantizador **106** para o codificador multiplexador de vídeo **170**, conforme indicado pela linha **125** na Figura 1. O codificador multiplexador de vídeo **170** ordena os coeficientes da transformada quantizados para cada bloco empregando um procedimento de varredura em ziguezague, convertendo deste modo o conjunto bidimensional de coeficientes de transformada quantizados em um conjunto unidimensional. Cada coeficiente quantizado com um valor não zero no conjunto unidimensional é então representado como um par de valores a que se refere como *uniforme* e *processado*, sendo *uniforme* o valor do coeficiente quantizado e *processado* o número dos coeficientes de valor zero consecutivos que precedem o coeficiente em questão. Os valores *processado* e *uniforme* são ainda comprimidos no codificador multiplexador de vídeo **170** empregando-se codificação de entropia, codificação de comprimento variável, por exemplo (VLC), ou codificação aritmética.

Uma vez os valores *processado* e *uniforme* codificados com entropia empregando-se um processo adequado, o codificador multiplexador de vídeo **170** combina ainda as mesmas informações de controle, também codificados com entropia empregando-se um processo adequado para o tipo de informações em questão, para formar uma única corrente de bits compactadas das informações de imagem codificadas **135**. Deve se observar que, embora a codificação com entropia tenha sido descrita em conexão com operações executadas pelo codificador multiplexador de vídeo **170**, pode ser provida em implementações alternativas uma unidade de codificação com entropia separada.

Uma versão localmente decodificada do macrobloco é também formada no codificador **100**. Isto se faz fazendo-se passar os coeficientes de transformação quantizados para cada bloco, emitidos pelo quantizador **106**, através do quantizador inverso **108** e aplicando-se uma transformada DCT inversa no bloco de transformação inversa **110**. Deste modo é construído um conjunto reconstruído de valores em pixels para cada bloco do macrobloco. Os dados de imagem decodificados resultantes são introduzidos no combinador **112**. No modo de Codificação-INTRA, o comutador **114** é ajustado de modo que a entrada no combinador **112** através do comutador **114** seja zero. Deste modo, a operação executada pelo combinador **112** é equivalente à passagem dos dados de imagem decodificados inalterados.

À medida que os macroblocos subseqüentes do quadro corrente são recebidos e são submetidos às etapas de codificação e decodificação local anteriormente descritas nos blocos **104**, **106**, **108**, **110** e **112**, uma versão decodificada do quadro INTRA-codificado é construída na memória de quadros **120**. Quando o último macrobloco do quadro atual tiver sido INTRA-codificado e subseqüentemente decodificado, a memória de quadros **120** contém um quadro completamente decodificado, disponível para uso como um quadro de referência de predição de movimento na codificação de um quadro de vídeo subseqüentemente recebido no formato INTER-codificado.

Será agora descrita a operação do codificador **100** no modo de Codificação-INTER. No modo de Codificação-INTER, o gerenciador de controle **160** opera o comutador **102** para receber a sua entrada da linha **117**, que compreende a saída do combinador **116**. O combinador **116** recebe o sinal de entrada de vídeo macrobloco a macrobloco da entrada **101**. À medida que o combinador **116** recebe os blocos de valores de luminância e croma que constituem o macrobloco, ele forma blocos correspondentes de informações de erro de predição. As informações de erro de predição representam a diferença entre o bloco em questão e a sua predição, produzida no bloco de predição de movimento compensado **150**. Mais especificamente, as informações de erro de predição para cada bloco do macrobloco compreendem um arranjo bidimensional de valores, cada um dos quais representa a diferença entre o valor em pixels no bloco de informações de luminância ou croma que estão sendo codificados, e um valor em pixels decodificado

obtido pela formação de uma predição de movimento compensado para o bloco, de acordo com o procedimento a ser descrito abaixo. Assim, no sistema de codificação de vídeo exemplar considerado aqui em que cada macrobloco compreende, por exemplo, um conjunto de blocos de 8 x 8 compreendendo valores de luminância e de crominância, as informações de erro de predição para cada bloco do macrobloco compreende do mesmo modo um arranjo de 8 x 8 de valores de erro de predição.

As informações de erro de predição para cada bloco do macrobloco são passadas ao bloco de transformação DCT **104**, que executa uma transformada de cossenos discretos bidimensional em cada bloco de valores de erro de predição para produzir um arranjo bidimensional de coeficientes da transformada DCT para cada bloco. O bloco de transformação DCT **104** produz um conjunto de valores de coeficientes para cada bloco de erro de predição, correspondendo o número de valores de coeficiente às dimensões dos blocos que constituem o macrobloco (neste caso 8 x 8). Os coeficientes da transformada derivados de cada bloco de erro de predição são passados para o quantizador **106** onde eles são quantizados empregando-se o parâmetro de quantização QP, de um modo análogo ao descrito acima em conexão com a operação do codificador no modo de codificação INTRA. Conforme já descrito, a seleção do parâmetro de quantização QP é controlada pelo gerenciador de controle **160** através da linha de controle **115**.

Os coeficientes DCT quantizados que representam as informações de erro de predição para cada bloco do macrobloco são passados do quantizador **106** para o codificador multiplexador de vídeo **170** conforme indicado pela linha **125** na Figura 1. Assim como no modo de Codificação-INTRA, o codificador multiplexador de vídeo **170** ordena os coeficientes de transformada para cada bloco de erro de predição empregando um procedimento de varredura em ziguezague determinado e em seguida representa cada coeficiente quantizado com valor não zero como um par *processado-uniforme*. Ele ainda comprime os pares *processado-uniforme* empregando uma codificação com entropia, de um modo análogo ao descrito acima em conexão com o modo de Codificação-INTRA. O codificador multiplexador de vídeo **170** também recebe informações de vetor de movimento (descritas abaixo) do bloco de codificação de campo de movimento **140** através da linha **126** e informações de controle do gerenciador de controle **160**. Ele codifica com entropia, as

informações de vetor de movimento e informações de controle e forma uma única corrente de bits de informações de imagem codificadas, compreendendo **135** o vetor de movimento codificado com entropia, informações de erro de predição e informações de controle.

Os coeficientes DCT quantizados que representam as informações de erro de predição para cada bloco do macrobloco são também passados do quantizador **106** ao quantizador inverso **108**. Aqui eles são quantizados inversamente e os blocos resultantes dos coeficientes DCT quantizados inversamente são aplicados ao bloco de transformada DCT inversa **110**, onde eles são submetidos à transformação DCT inversa para produzir blocos decodificados localmente de valores de erro de predição. Os blocos localmente decodificados de valores de erro de predição são então introduzidos no combinador **112**. No modo de Codificação-INTER, o comutador **114** é ajustado de modo que o combinador **112** também receba os valores previstos em pixels para cada bloco do macrobloco, gerados pelo bloco de predição de movimento compensado **150**. O combinador **112** combina cada um dos blocos decodificados localmente de valores de erro de predição com um bloco correspondente de valores em pixels previstos para produzir os blocos de imagem reconstruídos e os armazena na memória de quadros **120**.

À medida que os macroblocos subseqüentes do sinal de vídeo são recebidos da fonte de vídeo e são submetidos às etapas de codificação e decodificação já descritas nos blocos **104**, **106**, **108**, **110**, **112**, uma versão decodificada do quadro é construída na memória de quadros **120**. Quando o último macrobloco do quadro tiver sido processado, a memória de quadros **120** conterá um quadro completamente decodificado, disponível para uso como um quadro de referência de predição de movimento na codificação de um quadro de vídeo subseqüentemente recebido no formato INTER-codificado.

Serão agora considerados os detalhes da predição de movimento compensado executados pelo codificador de vídeo **100**.

Qualquer quadro codificado no formato de Codificação-INTER exige um quadro de referência para a predição de movimento compensado. Isto significa, necessariamente, que quando se codifica uma seqüência de vídeo, o primeiro quadro a ser codificado, quer ele seja o primeiro quadro na seqüência ou um outro quadro, deve ser codificado no formato de Codificação-INTRA. Isto por sua vez, significa que, quando o

codificador de vídeo **100** for comutado para o modo de Codificação-INTER pelo gerenciador de controle **160**, um quadro de referência completo, formado pela decodificação local de um quadro anteriormente codificado, já se encontra disponível na memória de quadros **120** do codificador. Geralmente, o quadro de referência é formado por decodificação local ou de um quadro INTRA-codificado ou um quadro INTER-codificado.

Na descrição abaixo será pressuposto que o codificador executa a predição de movimento compensado com base em macroblocos, isto é, um macrobloco é o menor elemento de um quadro de vídeo que pode ser associado com informações de movimento. Será ainda pressuposto que uma predição para um macrobloco dado é formada pela identificação de uma região de 16 x 16 valores no componente de luminância do quadro de referência que apresenta a melhor correspondência com 16 x 16 valores de luminância do macrobloco em questão. A predição de movimento compensado em um sistema de codificação de vídeo em que as informações de movimento podem ser associadas com elementos menores do que o macrobloco será considerada mais tarde no texto.

A primeira etapa na formação de uma predição para um macrobloco do quadro corrente é executada pelo bloco de estimativa de movimento **130**. O bloco de estimativa de movimento **130** recebe os blocos de valores de luminância e crominância que constituem o macrobloco corrente do quadro a ser codificado através da linha **128**. Ele então executa uma operação de emparelhamento de blocos a fim de identificar uma região no quadro de referência que apresenta a melhor correspondência com o macrobloco corrente. A fim de executar a operação de emparelhamento de blocos, o bloco de estimativa de movimento **130** acessa os dados do quadro de referência armazenados na memória de quadros **120** através da linha **127**. Mais especificamente o bloco de estimativa de movimento **130** executa o emparelhamento de blocos calculando os valores de diferença (somadas de diferenças absolutas, por exemplo) que representam a diferença em valores em pixels entre o macrobloco sendo examinado e as regiões em pixels candidatas com a melhor correspondência a partir de um quadro de referência armazenadas na memória de quadros **120**. O valor de diferença é produzido para regiões candidatas com todos os desvios possíveis dentro de uma região de busca predefinida do quadro de referência e o bloco de estimativa de movimento **130** determina o valor mínimo de diferença calculado. A região

candidata que resulta no menor valor de diferença é selecionada como a região com a melhor correspondência. O desvio do macrobloco corrente em relação à região com a melhor correspondência identificada no quadro de referência define um "vetor de movimento" para o macrobloco em questão. O vetor de movimento tipicamente compreende um par de números, um descrevendo o deslocamento horizontal (Δx) entre o macrobloco e a região com a melhor correspondência do quadro de referência, a outra representando o deslocamento vertical (Δy).

Tendo o bloco de estimativa de movimento **130** produzido um vetor de movimento para o macrobloco, ele passa o vetor de movimento ao bloco de codificação de campo de movimento **140**. O bloco de codificação de campo de movimento **140** aproxima o vetor de movimento recebido do bloco de estimativa de movimento **130** usando um modelo de movimento compreendendo um conjunto de funções de base e coeficientes de movimento. Mais especificamente o bloco de codificação de campo de movimento **140** representa o vetor de movimento como um conjunto de valores de coeficientes de movimento que, quando multiplicados pelas funções de base, formam uma aproximação do vetor de movimento. Tipicamente é usado um modelo de movimento de translação tendo somente dois coeficientes de movimento e funções de base, mas podem também ser usados modelos de movimento de complexidade maior.

Os coeficientes de movimento são passados do bloco de codificação de campo de movimento **140** ao bloco de predição de movimento compensado **150**. O bloco de predição de movimento compensado **150** também recebe a região com a melhor correspondência de valores em pixels identificada pelo bloco de estimativa de movimento **130** proveniente da memória de quadros **120**. Empregando a representação aproximada do vetor de movimento gerado pelo bloco de codificação de campo de movimento **140** e os valores em pixels da região de pixels com a melhor correspondência provenientes do quadro de referência, o bloco de predição de movimento compensado **150** terá um conjunto de valores em pixels previstos para cada bloco do macrobloco corrente. Cada bloco de valores em pixels previstos é passado ao combinador **116** onde os valores em pixels previstos são subtraídos dos valores em pixels reais (entrada) no bloco correspondente do macrobloco corrente. Deste modo obtém-se um conjunto de blocos de erro de predição para o

macrobloco.

Agora será descrita a operação do decodificador de vídeo **200**, mostrado na Figura 2. O decodificador **200** compreende um decodificador multiplexador de vídeo **270**, que recebe uma corrente de bits de vídeo codificada **135** proveniente do codificador **100** e desmultiplexadora-a nas suas partes constituintes, um quantizador inverso **210**, um transformador DCT inverso **220**, um bloco de predição de movimento compensado **240**, uma memória de quadros **250**, um combinador **230**, um gerenciador de controle **260** e uma saída **280**.

O gerenciador de controle **260** controla a operação do decodificador **200** em resposta ao quadro a ser decodificado consistir em um codificado por Codificação-INTRA ou por Codificação-INTER. Um sinal de controle de disparo INTRA/INTER, que faz com que o decodificador comute entre os modos de codificação, é derivado, por exemplo, das informações do tipo de imagem associada com cada quadro de vídeo comprimido recebido do codificador. O sinal de controle de disparo INTRA/INTER é extraído da corrente de bits de vídeo codificada pelo decodificador multiplexador de vídeo **270** e passada ao gerenciador de controle **260** através da linha **215**.

A decodificação de um quadro INTRA-codificado é executada numa base de macrobloco a macrobloco, sendo cada macrobloco decodificado substancialmente logo que as informações codificadas referentes a ele são recebidas na corrente de bits de vídeo **135**.

O decodificador multiplexador de vídeo **270** separa as informações codificadas para os blocos do macrobloco das informações de controle possíveis referentes ao macrobloco em questão. As informações codificadas para cada bloco de um macrobloco INTRA-codificado compreendem palavras de código de comprimento variável representando os valores *uniforme* e *processado* codificados com entropia para os coeficientes DCT quantizados diferentes de zero do bloco. O decodificador multiplexador de vídeo **270** decodifica as palavras de código de comprimento variável empregando um processo de decodificação de comprimento variável correspondente ao processo de codificação usado no codificador **100** e assim recupera os valores *uniforme* e *processado*. Ele então reconstruiu o conjunto de valores de coeficientes da transformada quantizados para cada bloco do macrobloco e os passa para o quantizador inverso **210**. Qualquer informação de controle que se refere ao



macrobloco é também decodificada no decodificador multiplexador de vídeo **270** empregando-se um processo de decodificação adequado e é passada ao gerenciador de controle **260**. Mais especificamente, as informações referentes ao nível de quantização aplicado aos coeficientes da transformada são extraídos da corrente de bits codificada pelo decodificador multiplexador de vídeo **270** e fornecidas ao gerenciador de controle **260** através da linha de controle **217**. O gerenciador de controle, por sua vez, transmite estas informações ao quantizador inverso **210** através da linha de controle **218**. O quantizador inverso **210** quantiza inversamente os coeficientes DCT quantizados para cada bloco do macrobloco de acordo com as informações de controle e fornece os coeficientes DCT agora quantizados inversos ao transformador DCT inverso **220**.

O transformador DCT inverso **220** executa uma transformada DCT inversa nos coeficientes DCT quantizados inversos para cada bloco do macrobloco para formar um bloco decodificado de informações de imagem compreendendo valores reconstruídos em pixels. Os valores reconstruídos em pixels para cada bloco do macrobloco são passados através do combinador **230** para a saída de vídeo **280** do decodificador, onde eles podem ser fornecidos, por exemplo, a um dispositivo de visualização (não mostrado). Os valores reconstruídos em pixels para cada bloco são também armazenados na memória de quadros **250**. Como a predição de movimento compensado não é usada na codificação/decodificação de macroblocos INTRA-codificados, o gerenciador de controle **260** controla o combinador **230** para passar cada bloco de valores em pixels no estado em que se encontram à saída de vídeo **280** e memória de quadros **250**. À medida que os macroblocos subseqüentes do quadro INTRA-codificado são decodificados e armazenados, um quadro decodificado é progressivamente montado na memória de quadros **250**, e assim se torna disponível para emprego como um quadro de referência para predição de movimento compensado em conexão com a decodificação de quadros INTER-codificados subseqüentemente recebidos.

Os quadros INTER-codificados são também decodificados macrobloco a macrobloco, sendo cada macrobloco INTER-codificado substancialmente decodificado, logo que as informações codificadas a ele referentes são recebidas na corrente de bits **135**. O decodificador multiplexador de vídeo **270** separa as informações de erro de predição codificadas para cada bloco de um macrobloco INTER-codificado das informações do vetor

de movimento codificadas e das informações de controle possíveis referentes ao macrobloco em questão. Conforme já explicado acima, as informações de erro de predição codificadas para cada bloco do macrobloco compreendem palavras de código de comprimento variável representando os valores *uniforme* e *processado* codificados com entropia para os

5 coeficientes da transformada quantizados diferentes de zero do bloco de erro de predição em questão. O decodificador multiplexador de vídeo **270** decodifica as palavras de código de comprimento variável empregando um processo de decodificação de comprimento variável correspondendo ao processo de codificação usado no codificador **100** e deste modo recupera os valores *uniforme* e *processado*. Ele então reconstrói um conjunto de valores de

10 coeficientes quantizados de transformada para cada bloco de erro de predição e os passa ao quantizador inverso **210**. As informações de controle que se referem ao macrobloco INTER-codificado são também decodificadas no decodificador multiplexador de vídeo **270** empregando-se um processo de decodificação adequado e são passadas ao gerenciador de controle **260**. As informações que se referem ao nível de quantização aplicado aos

15 coeficientes de transformada dos blocos de erro de predição são extraídas da corrente de bits codificados e fornecidos ao gerenciador de controle **260** através da linha de controle **217**. O gerenciador de controle, por sua vez transmite estas informações ao quantizador inverso **210** através da linha **218**. O quantizador inverso **210** quantiza ao inverso os coeficientes DCT quantizados representando as informações de erro de predição para cada bloco do

20 macrobloco de acordo com as informações de controle e fornece os coeficientes DCT agora quantizados inversos ao transformador DCT inverso **220**. Os coeficientes DCT quantizados inversos que representam as informações de erro de predição para cada bloco são então transformadas inversas no transformador DCT inverso **220** para produzir um conjunto de valores de erro de predição reconstruídos para cada bloco do macrobloco.

25 As informações de vetor de movimento codificadas associadas com o macrobloco são extraídas da corrente de bits de vídeo codificada **135** pelo decodificador multiplexador de vídeo **270** e são decodificadas. As informações de vetor de movimento decodificadas deste modo obtidas são passadas através da linha de controle **225** ao bloco de predição de movimento compensado **240**, que reconstrui um vetor de movimento para o

30 macrobloco empregando o mesmo modelo de movimento que foi usado para codificar o

macrobloco INTER-codificado no codificador **100**. O vetor de movimento reconstruído aproxima o vetor de movimento originalmente determinado pelo bloco de estimativa de movimento **130** do codificador. O bloco de predição de movimento compensado **240** do decodificador utiliza o vetor de movimento reconstruído para identificar a localização de uma região de pixels reconstruída em um quadro de referência de predição armazenado na memória de quadros **250**. O quadro de referência pode ser um quadro INTRA-codificado anteriormente decodificado, por exemplo, ou um quadro INTER-codificado anteriormente codificado. Qualquer que seja o caso a região de pixels indicada pelo vetor de movimento reconstruído é usada para formar uma predição para o macrobloco em questão. Mais especificamente, o bloco de predição de movimento compensado **240** forma um conjunto de valores em pixels para cada bloco do macrobloco copiando os valores em pixels correspondentes da região de pixels identificada pelo vetor de movimento. A predição, isto é, os blocos de valores em pixels derivados do quadro de referência, são passados do bloco de predição de movimento compensado **240** para o combinador **230** onde eles são combinados com as informações de erro de predição decodificadas. Na prática, os valores em pixels de cada bloco previsto são acrescentados aos valores de erro de predição correspondentes reconstruídos emitidos pelo transformador DCT inverso **220**. Deste modo um conjunto de valores em pixels reconstruídos para cada bloco do macrobloco é obtido. Os valores em pixels reconstruídos são passados para a saída de vídeo **280** no decodificador e são também armazenados na memória de quadros **250**. À medida que os macroblocos subsequentes do quadro INTER-codificado são decodificados e armazenados, um quadro decodificado é progressivamente montado na memória de quadros **250** e se torna assim disponível para uso como um quadro de referência para predição de movimento compensado das demais quadros INTER-codificados.

Conforme foi explicado acima, em um sistema de codificação de vídeo típico, a predição de movimento compensado é executada com base em macroblocos, de modo que um macrobloco é o menor elemento de um quadro de vídeo que pode ser associado a informações de movimento. No entanto, a recomendação de codificação de vídeo sendo atualmente desenvolvida pelo Joint Video Team (JVT) de ISO/IEC MPEG (Motion Pictures Expert Group) e ITU-T VCEG (Video Coding Experts Group) permite que

as informações de movimento sejam associadas a elementos menores do que um macrobloco. Na descrição abaixo e em todo o restante do texto, referir-se-á à versão deste padrão de codificação de vídeo descrito no documento por T. Weigland: "Joint Model Number 1", Doc. JVT-A003, Joint Video Team de ISO/IEC MPEG e ITU-T CEG, janeiro de 2002, sendo o documento integralmente incorporado ao presente documento a título de referência. Para fins de simplicidade, esta versão da recomendação será denominada "JM1 do codec de JVT".

De acordo com o JM1 do codec de JVT, os quadros de vídeo são divididos em macroblocos de 16 x 16 pixels e são codificados numa base de macrobloco a macrobloco. A codificação executada segue os princípios básicos descritos acima em conexão com o codificador e o decodificador de vídeo genéricos das Figuras 1 e 2. No entanto, de acordo com o JM1, a predição de movimento compensado dos macroblocos INTER-codificados é executada de um modo que difere do anteriormente descrito. Mais especificamente, a cada um dos macroblocos é atribuído um "modo de codificação" dependendo das características do macrobloco e do movimento na sequência de vídeo. Sete dos modos de codificação são baseados na divisão de um macrobloco a ser INTER-codificado em uma série de sub-blocos, compreendendo cada um deles N x M pixels, e na associação das informações de movimento a cada um dos N x M sub-blocos, e não somente ao macrobloco como um todo. Cada um dos esquemas possíveis para se dividir um macrobloco em N x M sub-blocos, provido pelo JM1 do codec de vídeo de JVT, é ilustrado na Figura 4 dos desenhos apensos. Conforme pode se ver da figura, as divisões possíveis são: 16 x 16, 8 x 16, 16 x 8, 8 x 8, 4 x 8, 8 x 4 e 4 x 4. Assim, se o modo de codificação atribuído a um macrobloco específico for, por exemplo, o modo 16 x 8, o macrobloco é dividido em dois sub-blocos tendo 16 x 8 pixels cada e os dois sub-blocos recebem as suas próprias informações de movimento. Além disso, é provido um oitavo modo de codificação, conhecido como modo DE PULO (ou de pulo). Se este modo for atribuído a um macrobloco, isto indica que o macrobloco deve ser copiado do quadro de vídeo de referência sem se empregar a predição de movimento compensado.

A decisão quanto à escolha do modo de codificação para um macrobloco dado é tipicamente tomada como parte do processo de estimativa de movimento. Mais



especificamente, em um codificador de vídeo tal como o ilustrado na Figura 1, mas implementado para permitir o emprego de diferentes modos de codificação de acordo com o JM1 do codec de JVT, o processo de estimativa de movimento executado pelo bloco de estimativa de movimento 130 é repetido para cada divisão possível do macrobloco nos $N \times M$ sub-blocos e para o modo de pulo. A estimativa de movimento para o modo de pulo é muito simples, uma vez que não é necessária nenhuma busca de movimento, mas um vetor constante com o valor zero é atribuído para este modo. Para os demais modos de Codificação-INTER, o movimento é tipicamente estimado executando-se uma operação de emparelhamento de blocos para cada bloco de movimento no interior do macrobloco. Depois destas operações, o modo que minimizar uma determinada função de custos é selecionado para o macrobloco. A função de custo tipicamente combina o erro de predição com o número de bits estimados como necessários para se codificar o macrobloco e deste modo mede-se a eficiência relativa de cada modo de codificação.

Como um codificador de vídeo que opera de acordo com o JM1 do codec de JVT atribui um modo de codificação específico a cada macrobloco que é INTER-codificado, é necessário que um decodificador de vídeo correspondente esteja ciente daquele modo de codificação a fim de que ele decodifique corretamente as informações recebidas referentes ao macrobloco em questão. Portanto, uma indicação do modo de codificação atribuído a cada macrobloco é provida na corrente de bits de vídeo transmitida do codificador de vídeo ao decodificador de vídeo. A fim de se minimizar a quantidade de dados necessários para indicar os modos de codificação, o modo de codificação para cada macrobloco é indicado empregando-se codificação de comprimento variável. As palavras de código indicando os modos de codificação são atribuídas de tal modo, que a palavra de código mais curta é usada para representar o modo de codificação com uma maior probabilidade estatística de ocorrer. O JM1 do codec de JVT usa um único conjunto de códigos denominado "Universal Variable Length Codes (Códigos Universais de Comprimento Variável)" (UVLC) para representar todos os elementos sintáticos (de dados) na corrente de bits de vídeo e portanto este conjunto de palavras de código é também usado para representar as informações de modo de codificação para os macroblocos INTER-codificados. As palavras de código UVLC usadas no JM1 podem ser escritas na seguinte



Índice de palavras de código	Palavra de código UVLC
0	1
1	001
2	011
3	00001
4	00011
5	01001
6	01011
7	0000001
8	0000011
9	0001001
10	0001011
11	0100001

12	0100011
13	0101001
14	0101011
15	000000001
..	...

Tabela 2:

As Primeiras 16 Palavras de Código UVLC do JM1

Geradas de Acordo com o Esquema Apresentado na Tabela 1

5 O JM1 do codec de JVT pressupõe que o modo de pulo é estatisticamente o mais provável modo de codificação para o macrobloco. O número de macroblocos de modo de pulo antes do seguinte macrobloco com um modo não-PULO é indicado por uma única palavra de código UVLC empregando-se a Tabela 2. Os modos de codificação restantes são representados pelas palavras de código UVLC conforme mostrado na Tabela 3 abaixo:

Índice de Palavra de códigos	Modo	Palavra de código UVLC
-	DE PULO	Codificado pelo Comprimento do Processado
0	16 x 16	1
1	16 x 8	001
2	8 x 16	011
3	8 x 8	00001
4	8 x 4	00011
5	4 x 8	01001
6	4 x 4	01011

Tabela 3: Modos de Codificação de Macroblocos de JM1

10 Conforme já descrito, os modos N x M na tabela acima indica o tamanho dos blocos de movimento.

Um problema com a abordagem adotada no JM1 do codec JVT é que a pressuposição de que o modo de pulo é sempre o mais provável não é válida. Se a sequência

de vídeo contiver movimento global (efeito panorâmico, zoom etc.), o modo de pulo na verdade nunca é usado. Nestes casos a eficiência da compressão é seriamente degradada, especialmente com taxas de bits mais baixas, uma vez que o codec é forçado a usar modos muito acima dos modos de codificação do macrobloco.

5 São conhecidas duas soluções da técnica anterior para se melhorar a eficiência de codificação da predição de movimento compensado na presença de movimento global. A primeira destas é conhecida como "compensação de movimento global" e é usada, por exemplo, nos padrões de codificação de vídeo ISO MPEG-4 e ITU-T H.263+. O segundo processo é descrito no documento intitulado "Global Motion Vector Coding" por 10 Shijun Sun e Shawmin Lei, Doc. VCEG-020, ITU-T Video Coding Experts Group (VCEG) Meeting, Pattaya, Tailândia 4-7 de dezembro de 2001, incorporado ao presente documento a título de referência. Os conceitos básicos destes dois processos serão agora sucintamente descritos.

Conforme descrito no Apenso P "Reference Picture Resampling" de 15 International Telecommunications Union ITU-T Recommendation H.263 "Video Coding for Low Bit Rate Communication", fevereiro de 1998, a idéia por trás da compensação do movimento global é a geração de um quadro de referência para a compensação de movimento que cancele os efeitos do movimento global. A fim de se fazer isso, são necessárias operações complexas no decodificador para deformar o quadro de referência em 20 uma forma mais utilizável. Além disso, as informações adicionais devem ser enviadas ao decodificador para guiar a construção de novos quadros de referência para a compensação do movimento global. Mais especificamente, o esquema de compensação de movimento global empregado no padrão de codificação de vídeo H.263 usa um processo de reamostragem para gerar uma versão deformada do quadro de referência para emprego na 25 predição de movimento compensado da imagem corrente. Esta versão deformada do quadro de referência pode incluir alterações no formato, tamanho e localização em relação à imagem corrente. O processo de re-amostragem é definido em termos de um mapeamento entre os quatro cantos da imagem corrente e os quatro cantos correspondentes do quadro de referência. Pressupondo-se que o componente de luminância do quadro corrente tenha uma 30 dimensão horizontal H e uma dimensão vertical V, o mapeamento é executado definindo-se

quatro vetores de movimento conceptual \underline{v}^{00} , \underline{v}^{H0} , \underline{v}^{0V} , e \underline{v}^{HV} , descrevendo cada vetor de movimento conceptual a maneira como se deve deslocar cada um dos quatro cantos do quadro corrente de um modo tal que se os mapeie sobre o canto correspondente do quadro de referência. Esta operação é ilustrada na Figura 5. A predição de movimento compensado para o macrobloco do quadro corrente é então executada empregando-se emparelhamento de blocos em relação ao quadro de referência deformado. Isto complica o processo de emparelhamento de blocos, uma vez que o valor de cada pixel do quadro de referência deformado usado no processo de emparelhamento de blocos deve ser gerado por mapeamento dos valores em pixels no quadro de referência original (não deformado) nas coordenadas do quadro de referência deformado. Isto é feito empregando-se uma interpolação bilinear, que é uma operação dispendiosa sob o ponto de vista da computação. O leitor deve consultar o Anexo P do padrão de codificação de vídeo H.263 para detalhes adicionais do processo de interpolação bilinear empregado para gerar os valores em pixels do quadro de referência deformado.

A codificação de vetor de movimento global, conforme descrita no documento VCEG-020, a que se refere acima, é uma versão simplificada de compensação de movimento global. O quadro de referência é usado tal como é, mas informações adicionais são transmitidas para descrever o movimento global e são usados modos de macroblocos adicionais para indicar quando são usados vetores de movimento global. Esta abordagem é menos complexa do que a técnica de compensação de movimento global que acabamos de descrever, mas há associada com ela uma complexidade adicional com o codificador. Mais exatamente, o codificador deve executar operações extra de estimativa de movimento para encontrar os parâmetros de movimento global e ela também precisa avaliar um número maior de modos de macrobloco para descobrir o ideal. Além disso, as informações de movimento global adicionais que precisam ser transmitidas se tornam demasiado numerosas para um vídeo de baixa resolução.

Levando-se em vista a discussão precedente, deve ser observado que há um problema técnico significativo que permanece sem solução no tocante à codificação de uma sequência de vídeo digital na presença de movimento global, tal como translação, efeito panorâmico ou zoom da câmara. Mais especificamente, cada uma das três soluções de

codificação de vídeo da técnica anterior descritos acima tem alguma forma de inconveniente técnico. O JM1 do codec JVT, por exemplo, não tem predição especial para levar em conta o movimento global nas seqüências de vídeo. Portanto, quando tal movimento ocorre, ele faz com que o codificador de vídeo selecione modos de codificação de macrobloco que modelam explicitamente o movimento. Isto leva a uma degradação significativa em eficiência de codificação, uma vez que o componente de movimento global é codificado em cada macrobloco INTER-codificado (ou sub-bloco). A técnica de compensação de movimento global (conforme dada pelo Apêndice P do padrão de codificação de vídeo H.263) leva o movimento global em conta deformando quadros de referência empregados na predição de movimento compensado e, portanto proporciona uma eficiência de codificação melhorada em comparação com o sistema em que nenhuma medida especial é tomada para codificar o movimento global. No entanto, o processo de deformação é complexo do ponto de vista da computação e deve-se transmitir informações adicionais na corrente de bits de vídeo codificada para permitir uma decodificação correta da seqüência de vídeo. Embora a técnica correlata de codificação de vetor de movimento global seja menos exigente do ponto de vista da computação do que a compensação de movimento global, ela realmente implica em um certo aumento na complexidade do codificador e continua havendo a necessidade de se ter que transmitir informações adicionais na corrente de bits de vídeo para permitir uma decodificação correta dos dados de vídeo.

Portanto, um objetivo da presente invenção consiste na combinação da simplicidade da compensação de movimento local com a eficiência de codificação da compensação de movimento global para produzir um sistema de codificação de vídeo com um desempenho de compressão significativamente melhorado e um aumento desprezível em complexidade.

Sumário da Invenção

A fim de se superar, ou pelo menos se mitigar até um ponto considerável, os problemas associados com a codificação de movimento global nos sistemas de codificação de vídeo da técnica anterior, a presente invenção é baseada em uma redefinição do conceito do modo de pulso empregado no JM1 do codec JVT. O processo de acordo com a invenção não somente proporciona um melhoramento na eficiência de codificação na presença de

movimento global (isto é, de movimento que afeta a área integral do quadro de vídeo), mas também permite que se represente de um modo eficiente o movimento regional.

De acordo com a invenção, o conceito do modo de pulo é redefinido de um modo tal que um macrobloco a que é atribuído o modo de pulo ou é associado com um vetor de movimento zero (não ativo), sendo neste caso tratado do mesmo modo como um macrobloco no modo de pulo convencional e copiado diretamente do quadro de referência, ou então ele é associado com um vetor de movimento não zero (ativo). A decisão de se associar um macrobloco a um vetor de movimento zero ou não zero é executada pela análise do movimento de outros macroblocos ou de sub-blocos em uma região circundando o macrobloco a ser codificado. Se for constatado que a região circundante apresenta um tipo determinado de movimento, é gerado um vetor de movimento não zero representativo daquele movimento e este vetor é associado ao macrobloco corrente. Mais especificamente, a continuidade, velocidade ou desvio de movimento nos macroblocos ou sub-blocos circundantes pode ser analisada. Se o movimento da região circundante apresentar, por exemplo, um determinado nível de continuidade, uma velocidade comum determinada ou uma forma específica de divergência um vetor de movimento representativo daquele movimento pode ser atribuído ao macrobloco corrente a ser codificado. Por outro lado, se a região que circunda o macrobloco corrente não apresentar tal continuidade, velocidade comum ou divergência e tiver um nível insignificante de movimento, ao macrobloco a ser codificado é atribuído um vetor de movimento zero, fazendo-se com que seja copiado diretamente do quadro de referência, exatamente do mesmo modo como se ele fosse um macrobloco no modo de PULO convencional. Deste modo, de acordo com a invenção, os macroblocos do modo DE PULO podem se adaptar ao movimento na região que os circunda, permitindo que se leve o movimento global ou regional em conta de um modo eficiente.

Em uma modalidade vantajosa da invenção, os macroblocos ou sub-blocos circundantes cujo movimento está sendo analisado são macroblocos anteriormente codificados vizinhos do macrobloco a ser codificado. Isto assegura que as informações de movimento referentes à região que circunda um macrobloco esteja disponível no codificador (decodificador) quando um macrobloco corrente estiver sendo codificado (decodificado) e

31
A

possa ser usado diretamente para se determinar o vetor de movimento a ser atribuído ao macrobloco corrente. Esta abordagem permite que a análise de movimento da região circundante executada no codificador seja duplicada exatamente no decodificador. Isto, por sua vez, significa que de acordo com a invenção, nenhuma informação adicional precisa ser enviada ao decodificador a fim de modelar o movimento global ou regional.

Conforme se tornará evidente da descrição detalhada da invenção apresentada abaixo, a redefinição da idéia do modo de pulo conforme proposto pela presente invenção tem vantagens técnicas significativas em comparação com os processos de codificação de vídeo da técnica anterior anteriormente descritos. Mais especificamente, o processo de acordo com a invenção permite que o movimento global e regional no interior da seqüência de vídeo seja levado em conta de um modo eficiente sem haver a necessidade de deformação complexa do quadro de referência ou de qualquer outra operação dispendiosa do ponto de vista da computação. Além disso, ao contrário do que ocorre tanto com o processo de codificação de compensação de movimento global como com o de vetor de movimento global já descritos, não há necessidade de nenhuma informação adicional ser transmitida na corrente de bits de vídeo para permitir uma decodificação correta dos dados de vídeo. Além disso, basta uma quantidade mínima de modificação para incorporar o processo de acordo com a invenção aos sistemas de codificação de vídeo existentes que empregam o conceito de macroblocos em modo de pulo.

Estas e outras características, aspectos e vantagens das modalidades da presente invenção se tornarão evidentes consultando-se a descrição detalhada abaixo em conjunto com os desenhos apensos. Deve ficar subentendido, no entanto, que os desenhos são destinados somente para o fim de ilustração e não como uma definição dos limites da invenção.

Descrição Sucinta dos Desenhos

A Figura 1 é um diagrama esquemático de blocos de um codificador genérico de vídeo de acordo com a técnica anterior.

A Figura 2 é um diagrama esquemático de blocos de um decodificador genérico de vídeo de acordo com a técnica anterior e correspondendo ao codificador mostrado na Figura 1.

informação adicional precisa ser transmitida do codificador ao decodificador.

A Figura 6 é um diagrama esquemático de blocos de um codificador de vídeo **600** implementado de acordo com uma modalidade da invenção. Quando se codifica os quadros de uma seqüência de vídeo digital, o codificador **600** opera de um modo análogo ao já descrito em conexão com o codificador de vídeo da técnica anterior da Figura 1 para gerar quadros de vídeo INTER-codificados e INTRA-codificados comprimidos. A estrutura do codificador de vídeo mostrado na Figura 6 é substancialmente idêntica à do codificador de vídeo da técnica anterior mostrado na Figura 1, com modificações adequadas na parte de estimativa de movimento necessária para implementar o processo de codificação de vídeo de acordo com a invenção. Todas as partes do codificador de vídeo que implementam funções e operam de um modo idêntico ao codificador de vídeo da técnica anterior já descrito são identificados com números de referência idênticos.

Como a presente invenção se refere à codificação de quadros de vídeo no formato INTER-codificado e mais especificamente a detalhes da predição de movimento compensado executada como parte do processo de Codificação-INTER, omitir-se-á a descrição do codificador **600** no modo de Codificação-INTRA e as seções abaixo se concentrarão nas operações executadas pelo codificador no modo de Codificação-INTER.

No modo de Codificação-INTER, o gerenciador de controle do codificador de vídeo **160** opera o comutador **102** para receber a sua entrada da linha **117**, que consiste na saída do combinador **116**. O combinador **116** recebe o sinal de entrada de vídeo macrobloco a macrobloco da entrada **101**. À medida que o combinador **116** recebe os blocos de valores de luminância e de croma que constituem o macrobloco, ele forma blocos correspondentes de informações de erro de predição, representando a diferença entre o bloco em questão e a sua predição, produzida no bloco de predição de movimento compensado **650**.

As informações de erro de predição para cada bloco do macrobloco são passadas para o bloco de transformação DCT **104**, que executa uma transformada de cossenos discretos bidimensional em cada bloco de valores de erro de predição para produzir um conjunto bidimensional de coeficientes de transformada DCT para cada bloco. Estes são passados ao quantizador **106** onde eles são quantizados empregando-se o

31

parâmetro de quantização QP. A seleção do parâmetro de quantização QP é controlado pelo gerenciador de controle **160** através da linha de controle **115**.

Os coeficientes DCT quantizados representando as informações de erro de predição para cada bloco do macrobloco são então passados do quantizador **106** para o codificador multiplexador de vídeo **170**, através da linha **125**. O codificador multiplexador de vídeo **170** ordena os coeficientes da transformada para cada bloco de erro de predição empregando um procedimento de varredura em ziguezague, representa cada coeficiente quantizado com um valor não zero como um par *processado-uniforme* e comprime os pares de *processado-uniforme* empregando a codificação com entropia. O codificador multiplexador de vídeo **170** também recebe as informações de vetor de movimento do bloco de codificação de campo de movimento **640** através da linha **126** e as informações de controle do gerenciador de controle **160**. Ele codifica com entropia as informações de vetor de movimento e as informações de controle e forma uma única corrente de bits de informações de imagem codificadas, compreendendo **135** o vetor de movimento codificado com entropia, erro de predição e informações de controle.

Os coeficientes DCT quantizados representando as informações de erro de predição para cada bloco do macrobloco são também passados do quantizador **106** ao quantizador inverso **108**. Aqui eles são quantizados inversos e os blocos resultantes dos coeficientes DCT quantizados inversos são aplicados ao bloco de transformada DCT inversa **110**, onde eles são submetidos a transformação DCT inversa para produzir blocos de valores de erro de predição decodificados localmente. Os blocos de valores de erro de predição localmente decodificados são então introduzidos no combinador **112**. No modo de Codificação-INTER, o comutador **114** é ajustado de modo tal que o combinador **112** também receba valores em pixels previstos para cada bloco do macrobloco, gerados pelo bloco de predição de movimento compensado **650**. O combinador **112** combina cada um dos blocos de valores de erro de predição localmente decodificados com um bloco de valores em pixels previstos correspondente para produzir blocos de imagem reconstruídos e os armazena na memória de quadros **120**.

À medida que os macroblocos subsequentes do sinal de vídeo são recebidos da fonte de vídeo e são submetidos às etapas de codificação e decodificação anteriormente

ab

descritas nos blocos **104**, **106**, **108**, **110**, **112**, uma versão do quadro é construída na memória de quadros **120**. Quando o último macrobloco do quadro tiver sido processado, a memória de quadros **120** conterá um quadro completamente decodificado, disponível para uso como um quadro de referência de predição de movimento na codificação de um quadro de vídeo subsequentemente recebido no forma de INTER-codificado.

Será agora descrita com detalhes a predição de movimento compensado executada pelo codificador de vídeo **600**.

O codificador **600** executa a predição de movimento compensado de um modo análogo ao já descrito para o codec JVT. Em outras palavras, ele é adaptado para atribuir um modo de codificação a cada macrobloco INTER-codificado dependendo das características do macrobloco e do movimento na sequência de vídeo que está sendo codificada. Quando se examina qual o modo de codificação que se deve atribuir a um macrobloco específico, o bloco de estimativa de movimento **630** executa uma operação de estimativa de movimento para cada modo de codificação de cada vez. O bloco de estimativa de movimento **630** recebe os blocos de valores de luminância e crominância que constituem o macrobloco a ser codificado para uso na estimativa de movimento através da linha **128** (veja Figura 6). Ele então seleciona cada um dos modos de codificação possíveis sucessivamente, um de cada vez, e executa a estimativa de movimento a fim de identificar a melhor correspondência para o macrobloco no quadro de referência, com base no modo de codificação selecionado e nos valores em pixels do macrobloco a ser codificado. (A melhor correspondência compreenderá uma ou mais regiões com a melhor correspondência de valores em pixels, dependendo do modo de codificação). Cada melhor correspondência é associada com um valor de custo total, uma combinação linear da soma de diferenças absolutas entre os valores em pixels no macrobloco que está sendo examinado e a região que tem a melhor correspondência no quadro de referência, por exemplo, e um número de bits necessário para codificar o modo e representar os vetores de movimento. Quando se obtém a melhor correspondência para cada modo de codificação, o bloco de estimativa de movimento **630** seleciona aquele modo de codificação que resulta no valor de custo geral menor como o modo de codificação para o macrobloco corrente.

De acordo com a invenção, os modos de codificação usados pelo codificador

48

600 correspondem aos propostos pelo JM1 do codec JVT (mostrado na Tabela 3), exceto que o modo de PULO é redefinido para permitir a representação do movimento global e regional. Mais especificamente, o modo de PULO é modificado de um modo tal, que um vetor de movimento zero (não ativo) e um vetor de movimento não zero (ativo) é associado com cada macrobloco no modo de pulo, dependendo das características do movimento nos segmentos de imagem que circundam o macrobloco em questão. No exposto abaixo, este tipo de vetor de movimento será denominado "vetor de movimento de modo de pulo".

Quando se examina o modo de pulo como parte do processo de estimativa de movimento descrito acima executado sucessivamente para cada modo de codificação, o codificador determina primeiro se deve ser usado um vetor de movimento de modo de pulo zero ou um não zero. Para tal fim, o codificador é disposto de modo a analisar o movimento dos segmentos de imagem (macroblocos e/ou sub-blocos, por exemplo) que circundam o macrobloco a ser codificado. Se ele determinar que a região circundante apresenta um determinado tipo de movimento, se, por exemplo, ele tiver características que indicam um movimento global ou regional, ele gera um vetor de movimento de modo de pulo com um valor não zero representativo do movimento. Por outro lado, se o codificador determinar que a região que circunda o macrobloco corrente não apresenta um movimento global ou regional, mas tem um nível insignificante de movimento, ele gerará um vetor de movimento de modo de pulo de valor zero. Isto é, se o codificador determinar que o movimento na região que circunda o macrobloco corrente tem uma característica global, a codificação de modo de pulo é adaptada para levar em conta tal fato (gerando um vetor de movimento de modo de pulo com um valor não zero associado representativo do movimento). Alternativamente, se nenhum tal movimento estiver presente, é gerado um vetor de movimento de valor zero fazendo com que o modo de pulo conforme modificado pela invenção opere de um modo convencional, isto é um vetor de movimento de modo de pulo com valor zero faz com que o macrobloco seja copiado diretamente do quadro de referência.

Tendo executado as operações de estimativa do movimento para cada um dos modos de codificação disponíveis, inclusive para o modo de pulo conforme modificado de acordo com a invenção, o codificador 600 determina qual o modo de codificação que resulta no valor de custo global menor e seleciona tal modo como o modo de codificação para o

macrobloco em questão. Uma indicação do modo de codificação finalmente selecionado, uma palavra de código de comprimento variável, por exemplo, selecionada do conjunto de palavras de código apresentadas na Tabela 3 é associada com o macrobloco e incluída na corrente de bits de vídeo **635**. Isto permite que um decodificador correspondente identifique o modo de codificação para o macrobloco e reconstrua de modo correto o macrobloco usando a forma correta da predição de movimento compensado.

A análise do movimento em uma região que circunda um macrobloco a ser codificado para determinar se deve ser usado um vetor de movimento de modo de pulo com valor zero ou não zero será agora descrita com mais detalhes fazendo-se referência à Figura 8 dos desenhos apensos. A Figura 8 ilustra os elementos funcionais do bloco de estimativa de movimento **630** associados à geração dos vetores de movimento do modo de pulo. Estes incluem memória de informações de movimento **801**, o bloco de análise de movimento circundante **802**, bloco de geração de parâmetros de movimento ativo **803**, e bloco de geração de parâmetros de movimento zero **804**.

A decisão de se gerar um vetor de movimento de modo de pulo com um valor zero ou um vetor de movimento de modo de pulo com um valor não zero é tomada pelo bloco de análise de movimento circundante **802**. A decisão é tomada analisando-se e classificando-se o movimento de macroblocos ou sub-blocos em uma região predefinida que circunda o macrobloco a ser codificado usando-se um esquema de análise predeterminado. A fim de se executar a análise, o bloco de análise de movimento circundante **802** recupera as informações de movimento que se referem aos macroblocos e/ou sub-blocos na região circundante da memória de informações de movimento **801**. Dependendo dos detalhes específicos da implementação, o bloco de análise de movimento circundante pode ser arranjado para analisar a continuidade, velocidade ou desvio do movimento nos macroblocos ou sub-blocos circundantes. Se o movimento na região circundante apresentar um determinado nível de continuidade uma determinada velocidade comum (conforme ilustrado na Figura 9, por exemplo), ou uma forma específica de divergência, por exemplo, isto pode sugerir que se encontra presente alguma forma de movimento global ou regional. Conseqüentemente o bloco de análise de movimento circundante conclui que um "movimento ativo" se encontra presente na região circundante e deve ser usado um vetor de

5

15

25

30

W
D

circundante determinar, por exemplo, que há um ou mais macroblocos vizinhos estacionários, emprega-se um vetor de movimento de modo de pulo de valor zero.

Em uma primeira modalidade preferida de invenção o bloco de análise de movimento circundante **802** classifica o movimento na região que circunda o macrobloco de acordo com o procedimento abaixo de três etapas. Em primeiro lugar, o bloco de análise de movimento circundante recupera as informações de movimento para os macroblocos ou sub-blocos que circundam o macrobloco a ser codificado (isto é, os macroblocos anteriormente codificados vizinhos do macrobloco a ser codificado, conforme mostrado na Figura 9) e gera uma predição de vetor de movimento média para o macrobloco. A predição de vetor de movimento média é formada, por exemplo, de um modo análogo ao usado na predição de vetor de movimento de acordo com o JM1 do codec JVT (veja T. Weigland: "Joint Model Number 1", Doc. JVT-A003, Joint Video Team de ISO/IEC MPEG e ITU-T VCEG, janeiro de 2002). Em seguida o bloco de análise de movimento circundante determina se qualquer um dos componentes de vetor de movimento resultante tem um valor absoluto superior a um determinado valor de limiar (meio pixel, por exemplo). Se esta condição for preenchida, o movimento é classificado como "movimento ativo", caso contrário ele é classificado como "movimento não ativo". Finalmente dependendo dos resultados da classificação, o bloco de análise de movimento circundante **802** envia uma indicação ou ao bloco de geração de parâmetros de movimento ativo **803** ou ao bloco de geração de parâmetros de movimento zero **804** para gerar os parâmetros de movimento de modo de pulo adequados.

A implementação do bloco de análise de movimento circundante de acordo com a primeira modalidade da invenção é especialmente vantajosa por dois motivos. Em primeiro lugar, em um codec vídeo típico, tal como o codec JVT, um preditor médio é usado para prever vetores de movimento de blocos de imagens quadrados. De acordo com a primeira modalidade preferida, este mesmo preditor é usado no bloco de análise de movimento circundante e no bloco de geração de parâmetros de movimento ativo para analisar o movimento na região que circunda um macrobloco a ser codificado e para gerar parâmetros de movimento para os macrobloco no modo de PULO. Deste modo a invenção pode ser implementada com um efeito mínimo sobre a complexidade de implementação

Será agora descrita a operação de um decodificador de vídeo **700** de acordo com uma modalidade da invenção, fazendo-se referência à Figura 7. A estrutura do decodificador de vídeo ilustrada na Figura 7 é substancialmente idêntica à do decodificador de vídeo da técnica anterior mostrado na Figura 2, com modificações apropriadas àquelas partes do decodificador que executam as operações de estimativa de movimento. Todas as partes do decodificador de vídeo que implementam funções e operam de um modo idêntico ao decodificador de vídeo da técnica anterior já descritos acima são identificados com números de referência idênticos. Pressupõe-se ainda que o decodificador de vídeo da Figura

5

10

20

25

30

inversos no transformador DCT inverso **220** para produzir um conjunto de valores de erro de predição reconstruídos para cada bloco do macrobloco.

Tanto as informações de indicação de modo de codificação como de vetor de movimento codificado (se houver) associadas com o macrobloco são decodificadas no decodificador multiplexador de vídeo, sendo então passadas por meio da linha de controle **225** ao bloco de predição de movimento compensado **740**. O bloco de predição de movimento compensado **740** usa a indicação do modo de codificação e as informações de vetor de movimento (se houver) para formar uma predição para o macrobloco em questão. Mais especificamente, o bloco de predição de movimento compensado **740** forma um conjunto de valores em pixels para cada bloco do macrobloco copiando os valores em pixels correspondentes de uma região (ou regiões) de pixels em um quadro de referência. A predição, isto é, os blocos de valores em pixels derivados do quadro de referência são passados do bloco de predição de movimento compensado **740** ao combinador **230** onde eles são combinados com as informações de erro de predição decodificadas (se houver). Deste modo é obtido um conjunto de valores em pixels reconstruídos para cada bloco do macrobloco.

Os valores em pixels reconstruídos são passados à saída de vídeo **280** do decodificador e também são armazenadas na memória de quadros **250**. Conseqüentemente, à medida que os macroblocos subseqüentes do quadro INTER-codificado são decodificados e armazenados, é montada progressivamente um quadro decodificado na memória de quadros **250** e assim se torna disponível para uso como um quadro de referência para a predição de movimento compensado de outros quadros INTER-codificados.

De acordo com a invenção, o bloco de predição de movimento compensado **740** do decodificador **700** compreende um bloco de memória de informações de movimento **801**, um bloco de análise de movimento circundante **802**, um bloco de geração de parâmetros de movimento ativo **803** e um bloco de geração de parâmetros de movimento zero **804** análogos aos providos no codificador **600**. Estes blocos funcionais são usados para se determinar se um macrobloco codificado no modo de pulo deve ser associado com um vetor de movimento de modo de pulo de valor zero ou de valor não zero. Mais especificamente, quando é determinado que um macrobloco a ser codificado tinha sido

codificado no modo de pulo, o bloco de análise de movimento circundante **802** analisa e classifica o movimento dos macroblocos sub-blocos anteriormente decodificados e/ou em uma região predefinida que circunda o macrobloco a ser decodificado de um modo que corresponde exatamente ao empregado no codificador **600**. Como resultado da análise, o

5 macrobloco em questão ou é associado com um vetor de movimento de modo de pulso de valor não zero ou um vetor de movimento de modo de pulso de valor zero. Este vetor de movimento é então usado para formar uma predição para o macrobloco. Se o macrobloco for associado com um vetor de movimento de modo de pulso de valor zero ele é reconstruído por simples cópia de valores em pixels de uma localização correspondente no quadro de

10 referência. Se, por outro lado, ele for associado com um vetor de movimento de valor não zero, uma região de valores em pixels indicada pelo vetor de movimento não zero é usada para gerar os valores em pixels para o macrobloco.

Deve se observar que modificando-se o conceito do modo de pulso da maneira proposta pela invenção e executando-se a análise de movimento circundante no

15 decodificador, é possível se levar em conta o movimento global ou regional em uma sequência de vídeo sem haver a necessidade de transmitir informações explícitas sobre tal movimento na corrente de bits de vídeo.

Será agora descrito um dispositivo terminal compreendendo equipamento de codificação e decodificação de vídeo que pode ser adaptado para operar de acordo com a

20 presente invenção. A Figura 10 dos desenhos apensos ilustra um terminal de multimídia **80** implementado de acordo com a recomendação H.324 do ITU-T. O terminal pode ser considerado como um dispositivo transceptor de multimídia. Ele inclui elementos que capturam, codificam e multiplexam as correntes de dados de multimídia para transmissão através de uma rede de comunicações, assim como elementos que recebe,

25 desmultiplexadora, decodifica e apresenta o conteúdo de multimídia recebido. A recomendação H.324 de ITU-T define a operação global do terminal e se refere a outras recomendações que governam a operação das suas diversas partes constituintes. Este tipo de terminal de multimídia pode ser usado em aplicações de tempo real tais como em vídeo telefonia de conversação, ou em aplicações não no tempo real tais como a recuperação e/ou

30 a conversão em corrente de clipes de vídeo, proveniente de um servidor de conteúdo de

multimídia na Internet, por exemplo.

Dentro do contexto da presente invenção, deve ser observado que o terminal de acordo com H.324 mostrado na Figura 10 é somente um de uma série de implementações alternativas de terminal de multimídia adaptadas à aplicação do processo da presente invenção. Deve-se também observar que existe uma série de alternativas referentes à localização e implementação do equipamento de terminal. Conforme ilustrado na Figura 10, o terminal de multimídia pode estar localizado em um equipamento de comunicações conectado a uma rede de linha telefônica fixa tal como uma RTCP (Rede de Telefonia Comutada Pública) analógica. Neste caso o terminal de multimídia é equipado com um modem **91**, compatível com as recomendações V.8, V.34 e opcionalmente V.8bis de ITU-T. Alternativamente, o terminal de multimídia pode ser conectado a um modem externo. O modem permite a conversão dos dados digitais multiplexados e os sinais de controle produzidos pelo terminal de multimídia em uma forma analógica adequada para a transmissão através de RTCP. Ele ainda permite que o terminal de multimídia receba os dados e os sinais de controle em forma analógica da RTCP e converta-os em uma corrente de dados digitais que podem ser desmultiplexados e processados em um modo adequado pelo terminal.

Um terminal de multimídia H.324 pode também ser implementado de um modo tal que ele possa ser conectado diretamente a uma rede de linha fixa digital tal como a RDSI (Rede Digital de Serviços Integrados). Neste caso o modem **91** é substituído por uma interface de rede RDSI de usuário. Na Figura 10, esta interface de rede RDSI de usuário é representada pelo bloco alternativo **92**.

Os terminais de multimídia H.324 podem também ser adaptados para uso em aplicações de comunicação móvel. Se forem usados com uma conexão de comunicação sem fio, o modem **91** pode ser substituído com qualquer interface apropriada de sem fio, conforme representado pelo bloco alternativo **93** na Figura 10. Um terminal de multimídia H.324/M, por exemplo, pode incluir um transceptor de rádio que permite a conexão à rede de telefones móveis G/M2 atual de 2a. geração, ou ao UMTS (Sistema de Telefonia Móvel Universal/ Universal Mobile Telephone System) de 3a. geração proposto.

Deve ser observado que em terminais de multimídia projetados para a

5 denominado "codec".

10

20

30

de áudio compreende um par de codificador/decodificador. Ele converte os dados de áudio capturados pelo equipamento de áudio do terminal em uma forma adequada para transmissão através da conexão de comunicação e transforma os dados de áudio codificados recebidos da rede de volta em uma forma adequada para a reprodução, no alto-falante do terminal, por exemplo. A saída do codec de áudio é passada para um bloco de retardo 87. Este compensa os atrasos introduzidos pelo processo de codificação de vídeo e assegura assim a sincronização do conteúdo de áudio e de vídeo.

O bloco de controle de sistema 84 do terminal de multimídia controla a sinalização final para rede empregando um protocolo de controle adequado (bloco de sinalização 88) para estabelecer um modo comum de operação entre um terminal de transmissão e um de recepção. O bloco de sinalização 88 troca informações sobre as capacidades de codificação e decodificação dos terminais de transmissão e de recepção e pode ser usado para permitir que diversos modos de codificação do codificador de vídeo. O bloco de controle do sistema 84 também controla o uso da cifragem de dados. As informações referentes ao tipo de cifragem a ser usado na transmissão de dados são passadas do bloco de cifragem 89 ao multiplexador/demultiplexador (unidade MUX/DMUX) 90.

Durante a transmissão de dados do terminal de multimídia a unidade MUX/DMUX 90 combina as correntes de vídeo e de áudio codificadas e sincronizadas com entrada de dados provenientes do equipamento telemático 83 e possíveis dados de controle, para formar uma única corrente de bits. As informações referentes ao tipo de cifragem de dados (se houver) a ser aplicado à corrente de bits, providas pelo bloco de cifragem 89 são usadas para selecionar um modo de cifragem. De modo correspondente, quando um multiplexador e possivelmente a corrente de bits de multimídia possivelmente codificada estiver sendo recebido, a unidade MUX/DMUX 90 é responsável pela decifragem da corrente de bits, dividindo-a nos seus componentes de multimídia constituintes e passando esses componentes ao(s) decodificador(es)/decodificador(es) e/ou equipamento de terminal para a decodificação e reprodução.

Se o terminal de multimídia 80 for um terminal móvel, isto é, se ele for equipado com um transceptor de rádio 93, deve ser subentendido pelos peritos na técnica

48
a

que ele pode também compreender elementos adicionais. Em uma modalidade ele compreende uma interface de usuário que tem um monitor e um teclado, o que permite a operação do terminal de multimídia **80** por um usuário, uma unidade de processamento central, tal como um microprocessador, que controla os blocos responsáveis por diferentes

5 funções do terminal de multimídia, uma memória de acesso aleatório RAM, uma memória de leitura somente ROM e uma câmara digital. As instruções operacionais do microprocessador, isto o código de programa que corresponde às funções básicas do terminal de multimídia **80**, são armazenadas na memória de leitura somente ROM e podem ser executadas quando exigidas pelo microprocessador, sob controle do usuário, por

10 exemplo. De acordo como código de programa, o microprocessador usa o transceptor de rádio **93** para formar uma conexão com uma rede de comunicação móvel, permitindo que o terminal de multimídia **80** transmita informações para a rede de comunicação móvel através de um trajeto de rádio e receba informações dela.

O microprocessador monitora o estado da interface de usuário e controla a

15 câmara digital. Em resposta a um comando do usuário, o microprocessador instrui a câmara para registrar imagens digitais em RAM. Uma vez capturada uma imagem ou uma seqüência de vídeo digital, ou alternativamente durante o processo de captura, o microprocessador segmenta a imagem em segmentos de imagem (macroblocos, por exemplo) e usa o codificador para executar a codificação de movimento compensado dos

20 segmentos a fim de gerar uma seqüência de imagem comprimida, conforme explicada na descrição acima. Um usuário pode comandar o terminal de multimídia **80** para apresentar as imagens capturadas no seu monitor ou para enviar a seqüência de vídeo comprimida empregando um transceptor de rádio **93** a um outro terminal de multimídia, um vídeo telefone conectado a uma rede de linha fixa (RTCP – Rede de Telefonia Comutada Pública)

25 ou a algum outro dispositivo de telecomunicação. Em uma modalidade preferida, a transmissão de dados de imagem é iniciada logo que o primeiro segmento é codificado de modo que o receptor possa iniciar um processo de decodificação correspondente com um mínimo de atraso.

Embora tenha sido descrito dentro do contexto de modalidades específicas,

30 será evidente aos peritos na técnica que uma série de modificações e diversas alterações a

estas instruções possa ocorrer. Assim, embora a invenção tenha sido especificamente mostrada e descrita em relação a uma ou mais modalidades preferidas dela, deve ser subentendido pelos peritos na técnica que determinadas modificações ou alterações possam ser introduzidas nelas, sem que haja desvio do âmbito e espírito da invenção conforme apresentada acima.

Especificamente, de acordo com uma modalidade alternativa da invenção, o bloco de análise de movimento circundante **802** é adaptado para classificar o movimento de uma região circundante em mais de duas classes de movimento. Uma classificação significativa abrangendo três classes de movimento, por exemplo, seria "movimento contínuo", "movimento ativo" e "movimento não ativo". Deste modo podem ser gerados parâmetros de movimento especiais para o caso típico de movimento contínuo.

Em uma outra modalidade alternativa da invenção, o bloco de análise de movimento circundante é removido e o bloco de geração de parâmetros de movimento ativo é ativado para todos os macrobloco em modo de pulo.

De acordo com uma outra modalidade alternativa em vez de se empregar o bloco de análise de movimento circundante para indicar as informações de classificação, esta indicação é provida por outros meios (como uma informação colateral a nível de macrobloco, de fração, de imagem ou de sequência, por exemplo).

Em uma outra modalidade alternativa ainda, o bloco de análise de movimento circundante pode ser temporariamente desabilitado ou habilitado com tais meios.

Em uma outra implementação alternativa, a tabela de modos de macroblocos é novamente excluída dependendo da saída do bloco de análise de movimento circundante para se fornecer uma prioridade superior aos modos mais prováveis. Em uma modalidade correlata, a tabela de modos de macroblocos é completamente regenerada, dependendo da saída do bloco de análise de movimento circundante, por remoção do pulo, por exemplo.

Deve ser também observado que o bloco de geração de parâmetros de movimento **803** pode ser implementado de diversas maneiras. Em modalidades específicas da invenção ele é adaptado para gerar parâmetros de movimento, com base no caráter contínuo, velocidade ou desvio do movimento circundante. As informações colaterais adicionais podem também ser enviadas para guiar a geração dos parâmetros de movimento.

Em uma modalidade alternativa, o bloco de geração de parâmetros de movimento ativo é adaptado para emitir vetores de movimento múltiplos a serem usados em diferentes partes do macrobloco.

A invenção pode também ser usada para gerar predição de movimento compensado para outros modos de Macrobloco além do modo DE PULO ou em vez dele. Ela tampouco é limitada pela estrutura de macroblocos, mas pode ser usada em qualquer sistema de codificação de vídeo baseada em segmentação.

REIVINDICAÇÕES

1. Um método de predição de movimento compensado para uso na codificação de uma sequência de vídeo digital, a sequência de vídeo digital compreendendo um número de quadros, o método é caracterizado por:

atribuir um modo de codificação para um segmento do primeiro quadro da sequência a ser codificado usando a predição de movimento compensado em relação ao segundo quadro da sequência, o modo de codificação sendo um de um grupo de modos de codificação possíveis incluindo o modo de pulo,

realizar uma análise do movimento em uma região do primeiro quadro que circunda o segmento a ser codificado para determinar a característica de movimento na região;

onde o modo de pulo, quando atribuído ao segmento do primeiro quadro, indica, com base nas características do movimento da dita região; ou

(i) que o segmento tem um nível insignificante de movimento, neste caso, formando uma predição para o segmento ao copiar diretamente da região correspondente do segundo quadro, ou

(ii) que o segmento tem uma característica de movimento do movimento global ou regional na sequência de vídeo, neste caso, formando uma predição para o segmento de predição de movimento compensado em relação ao segundo quadro usando a informação de movimento representativa do movimento global ou regional.

2. Método de acordo com a reivindicação 1, caracterizado por a região do primeiro quadro que circunda o segmento a ser codificado compreender segmentos codificados previamente do primeiro quadro.

3. Método de acordo com a reivindicação 1, caracterizado por se for determinado que o movimento na região do primeiro quadro que circunda o segmento a ser codificado tem um nível insignificante, o modo de pulo é atribuído como o modo de codificação do segmento e a informação de movimento indicativa de movimento zero é associada com o segmento.

4. Método de acordo com a reivindicação 1, caracterizado por se for determinado que o movimento na região do primeiro quadro que circunda o segmento a ser codificado é característico do movimento global ou regional, o modo de pulo é atribuído como o modo de codificação do segmento e a informação de movimento indicativa de movimento global ou regional na região vizinha é associada com o segmento.

5. Método de acordo com a reivindicação 1, caracterizado por a análise compreende o exame da região do primeiro quadro que circunda o segmento a ser codificado para identificar pelo menos um dos tipos seguintes de movimento: movimento contínuo, movimento tendo uma velocidade comum, e movimento que tem um determinado desvio.

6. Método de acordo com a reivindicação 1, caracterizado por se um segmento do primeiro quadro na região do primeiro quadro que circunda o segmento a ser codificado não tiver nenhum movimento, o modo de pulo é atribuído como o modo de

codificação do segmento e a informação de movimento indicativa de movimento zero é associada com o segmento.

7. Método de predição de movimento compensado para uso na decodificação de uma sequência de vídeo digital codificada, a sequência de vídeo digital compreendendo um número de quadros, o método é caracterizado por compreender:

- receber uma indicação de um modo de codificação atribuído a um segmento de um primeiro quadro da sequência decodificada empregando-se a predição de movimento compensado em relação a um segundo quadro da sequência, sendo o modo de codificação um de um conjunto de modos de codificação possíveis incluindo um modo de pulo; e

- realizar uma análise do movimento em uma região do primeiro quadro que circunda o segmento a ser decodificado para determinar a característica de movimento na região;

- determinar, quando for indicado o modo de pulo como o modo de codificação atribuído ao segmento do primeiro quadro, baseado na característica do movimento na região,

- (i) se tal modo de pulo indica que o segmento tem um nível insignificante de movimento, sendo, em tal caso, a predição de movimento compensado para o segmento ser executado copiando-se diretamente de uma região correspondente do segundo quadro, ou

- (ii) se tal modo de pulo indica que o segmento tem uma característica de movimento do movimento global ou regional na sequência de vídeo, sendo em tal caso a predição de movimento compensado efetuada em relação ao segundo quadro empregando-se informações de movimento que representa tal movimento global ou regional.

8. Método de acordo com a reivindicação 7, caracterizado por a etapa de determinação compreender:

- a execução de uma análise de movimento em uma região anteriormente decodificada do primeiro quadro circundando o segmento a ser decodificado a fim de se determinar uma característica do movimento na região.

9. Método de acordo com a reivindicação 8, caracterizado por a análise compreender o exame da região anteriormente decodificada do primeiro quadro que circunda o segmento a ser decodificado a fim de se identificar pelo menos um dos seguintes tipos de movimento: movimento contínuo, movimento tendo uma velocidade comum, e movimento tendo um determinado desvio, ficando determinado, no caso do pelo menos um tipo de movimento ter sido identificado na região circundante, que o modo de pulso indica que o segmento a ser decodificado tem características de movimento do movimento global ou regional na sequência de vídeo.

10. Método de acordo com a reivindicação 8, caracterizado por se um segmento do primeiro quadro na região anteriormente decodificada do primeiro quadro circunda o segmento a ser codificado não tiver nenhum movimento, fica determinado que o modo de pulso indica que o segmento a ser decodificado tem um nível insignificante de movimento.

11. Codificador de vídeo disposto de modo a codificar a sequência de vídeo digital empregando-se predição de movimento compensado, a sequência de vídeo digital compreendendo um número de quadro, caracterizado por compreender:

- dispositivo para atribuir um modo de codificação a um segmento de um primeiro quadro da sequência a ser

codificado empregando-se predição de movimento compensado em relação a um segundo quadro da sequência, consistindo o modo de codificação em um de um conjunto de modos possíveis inclusive um modo de pulo; e

- dispositivo para realizar uma análise de movimento em uma região do primeiro quadro que circunda o segmento a ser codificado para determinar uma característica de movimento na região;

- dispositivo para a formação de uma predição para o segmento, indicando o modo de pulo, quando atribuído ao segmento do primeiro quadro que indica com base na característica de movimento na região, ou

- (i) que o segmento tem um nível insignificante de movimento, formando, em tal caso, os meios de formação uma predição para o segmento copiando diretamente de uma região correspondente do segundo quadro, ou

- (ii) que o segmento tem características de movimento de movimentos global ou regional na sequência de vídeo, formando, neste caso, os meios de formação uma predição para o segmento em relação ao segundo quadro empregando informações de movimento representativas do movimento global ou regional.

12. Codificador de acordo com a reivindicação 11, caracterizado por os meios de análise estarem dispostos de modo a analisar o movimento nos segmentos anteriormente codificados do primeiro quadro.

13. Codificador de acordo com a reivindicação 11, caracterizado por se os meios de análise determinarem que o movimento na região possuem um nível insignificante, o codificador é disposto de modo a atribuir o modo de pulo

como o modo de codificação do segmento e a associar as informações de movimento indicadoras de movimento zero com o segmento a ser codificado.

14. Codificador de acordo com a reivindicação 11, caracterizado por os meios de análise determinam que o movimento na região do primeiro quadro que circunda o segmento a ser codificado é característico de movimento global ou regional, o codificador é disposto de modo atribuir o modo de pulo como o modo de codificação do segmento e a associar as informações de movimento indicadoras do movimento global ou regional na região circundante com o segmento a ser codificado.

15. Codificador de acordo com a reivindicação 11, caracterizado por os meios de análise compreendem meios para a identificação de pelo menos um dos seguintes tipos de movimento: movimento contínuo, movimento tendo uma velocidade comum, movimento tendo um determinado desvio.

16. Codificador de acordo com a reivindicação 11, caracterizado por os meios de análise determinarem que um segmento do primeiro quadro na região do primeiro quadro que circunda o segmento a ser codificado não tem nenhum movimento, o codificador ser disposto de modo a atribuir o modo de pulo como o modo de codificação do segmento e a associar as informações de movimento indicadoras de movimento zero com o segmento a ser codificado.

17. Decodificador de vídeo disposto de modo a decodificar uma sequência de vídeo digital codificada empregando a predição de movimento compensado, a sequência de vídeo

digital compreendendo uma série de quadros, caracterizado por compreender:

- dispositivo para receber uma indicação de um modo de codificação atribuído a um segmento de um primeiro quadro da sequência codificada empregando a predição de movimento compensado em relação a um segundo quadro da sequência, consistindo o modo de codificação em um de um conjunto de modos de codificação possíveis inclusive um modo de pulo; e

- dispositivo para realizar análises de movimento em uma região do primeiro quadro que circunda o segmento a ser codificado para determinar uma característica do movimento na região;

- dispositivo para formar uma predição para o segmento, sendo que, quando o modo de pulo for indicado como o modo de codificação atribuído ao segmento do primeiro quadro, o decodificador é disposto de modo a determinar baseado na característica do movimento na região,

- (i) se o modo de pulo indica que o segmento tem um nível insignificante de movimento, formando neste caso os meios de formação a predição para o segmento copiando diretamente de uma região correspondente do segundo quadro, ou

- (ii) se o modo de pulo indica que o segmento tem característica de movimento de movimento global ou regional na sequência de vídeo, formado neste caso os meios de formação a predição para o segmento por predição de movimento compensado em relação ao segundo quadro empregando informações de movimento representativas de movimento global ou regional.

18. Decodificador de acordo com a reivindicação 17, caracterizado por compreender ainda meios de análise de movimento em uma região anteriormente decodificada do primeiro quadro que circunda o segmento a ser decodificado, para determinar uma característica do movimento na região.

19. Decodificador de acordo com a reivindicação 18, caracterizado por se os meios de análise determinam que o movimento na região possuem um nível insignificante, o codificador é disposto de modo a associar as informações de movimento indicadoras de movimento zero com o segmento a ser decodificado.

20. Decodificador de acordo com a reivindicação 18, caracterizado por se os meios de análise determinam que o movimento na região é característico do movimento global ou regional, o decodificador e disposto de modo associar as informações de movimento indicadoras de movimento global ou regional na região circundante com o segmento a ser decodificado.

21. Decodificador de acordo com a reivindicação 18, caracterizado por os meios de análise compreendem meios para a identificação de pelo menos um dos seguintes tipos de movimento: movimento contínuo, movimento que tem uma velocidade comum, movimento tendo um certo desvio.

22. Decodificador de acordo com a reivindicação 18, caracterizado por se os meios de análise determinam que um segmento anteriormente decodificado do primeiro quadro na região não tem nenhum movimento, o decodificador é disposto de modo a associar as informações de movimento indicadoras de movimento zero com o segmento a ser decodificado.

23. Um terminal multimídia, caracterizado por compreender:

- dispositivo para se obter uma sequência de vídeo digital; e

- um codificador de vídeo disposto de modo a codificar a sequência de vídeo digital empregando predição de movimento compensado, compreendendo a sequência de vídeo digital uma série de quadros, o codificador de vídeo compreende:

- dispositivo para atribuir um modo de codificação a um segmento de um primeiro quadro da sequência a ser codificada empregando predição de movimento compensando em relação a um segundo quadro da sequência, consistindo o modo de codificação em um de um conjunto de modos possíveis incluindo um modo de pulo; e

- dispositivo para realização de análises de movimento em uma região do primeiro quadro que circunda o segmento a ser codificado para determinar uma característica de movimento na região;

- dispositivos para a formação de uma predição para o segmento, indicando o modo de pulo, quando atribuído ao segmento do primeiro quadro, que indica baseado na característica de movimento na região; ou

- (i) que o segmento tem um nível insignificante de movimento, sendo em tal caso, o codificador de vídeo disposto de modo tal que os meios de formação formam uma predição para o segmento copiando diretamente de uma região correspondente do segundo quadro, ou

- (ii) que o segmento tem característica de movimento de movimento global ou regional na

sequência de vídeo, sendo neste caso o codificador de vídeo disposto de modo tal que os meios de formação formam uma predição para o segmento em relação ao segundo quadro empregando informações de movimento representativas do movimento global ou regional.

24. Terminal de multimídia caracterizado por:

- dispositivo para a aquisição de uma sequência de vídeo digital codificada; e
- um decodificador de vídeo disposto de modo a decodificar a sequência de vídeo digital codificada empregando predição de movimento compensado, compreendendo a sequência de vídeo digital uma série de quadros, o decodificador compreende:
 - dispositivo para receber uma indicação de um modo de codificação atribuído a um segmento de um primeiro quadro da sequência codificada empregando predição de movimento compensado em relação a um segundo quadro de sequência, sendo o modo de codificação um de um conjunto de modos de codificação possíveis que inclui um modo de pulo; e
 - dispositivo para realizar uma análise de movimento na região do primeiro quadro que circunda o segmento a ser codificado de modo a determinar uma característica de movimento na região;
 - dispositivos para a formação de uma predição para o segmento, sendo que, quando o modo de pulo for indicado como o modo de codificação atribuído ao segmento do primeiro quadro, o decodificador é disposto de modo a determinar, com base na característica de movimento na região

(i) se o modo de pulo indica que o segmento tem um nível insignificante de movimento, formando neste caso os meios de formação a predição para o segmento copiando diretamente de uma região correspondente do segundo quadro; ou

(ii) se o modo de pulo indica que o segmento tem característica de movimento de movimento global ou regional na sequência de vídeo, formando neste caso os meios de formação a predição para o segmento por predição de movimento compensado em relação ao segundo quadro usando as informações de movimento representativas do movimento global ou regional.

25. Codec de vídeo caracterizado por

- um codificador de vídeo disposto de modo a codificar uma sequência de vídeo digital usando predição de movimento compensado, compreendendo a sequência de vídeo digital uma série de quadros, compreendendo o codificador:

- dispositivos para atribuir um modo de codificação a um segmento de um primeiro quadro da sequência a ser codificada empregando predição de movimento compensado em relação a um segundo quadro da sequência, consistindo o modo de codificação em um de um conjunto de modos possíveis incluindo um modo de pulo; e

- dispositivo para realizar uma análise de movimento em uma região do primeiro quadro que circunda o segmento a ser decodificado para determinar uma característica de movimento na região;

- dispositivo para a formação de uma predição para o segmento, em que o modo de pulo, quando atribuído ao segmento do primeiro quadro, que indica com base na característica de movimento na região ou

(i) que o segmento tem um nível insignificante de movimento, sendo neste caso o codificador de vídeo de tal modo disposto que os meios formadores formam uma predição para o segmento copiando diretamente de uma região correspondente do segundo quadro, ou

(ii) que o segmento tem característica de movimento de movimento global ou regional na sequência de vídeo, sendo neste caso o codificador de vídeo disposto de tal modo que os meios de formação formam uma predição para o segmento em relação ao segundo quadro empregando as informações de movimento representativas do movimento global ou regional; e

- um decodificador disposto de modo a decodificar uma sequência de vídeo digital codificada empregando predição de movimento compensado, compreendendo a sequência de vídeo digital uma série de quadros, o decodificador compreendendo:

- dispositivo para receber uma indicação de um modo de codificação atribuído a um segmento de um primeiro quadro da sequência codificada usando predição de movimento compensado em relação a um segundo quadro da sequência, sendo o modo de codificação um de um conjunto de modos de codificação possíveis incluindo um modo de pulo; e

- dispositivo para realizar uma análise de movimento em uma região do primeiro quadro que circunda o segmento a ser codificado para determinar uma característica de movimento na região;

- dispositivo para a formação de uma predição para o segmento, sendo que, quando é indicado o modo de pulo como o modo de codificação atribuído ao segmento do primeiro quadro, o decodificador é

disposto de modo a determinar, com base em características de movimento na região;

(i) se o modo de pulo indica que o segmento tem um nível insignificante de movimento, formando neste caso os meios de formação a predição para o segmento copiando diretamente de uma região correspondente do segundo quadro, ou

(ii) se o modo de pulo indica que o segmento tem características de movimento de movimento global ou regional na sequência de vídeo, formando neste caso os meios de formação a predição para o segmento por predição de movimento compensado em relação ao segundo quadro empregando informações de movimento representativas do movimento global ou regional.

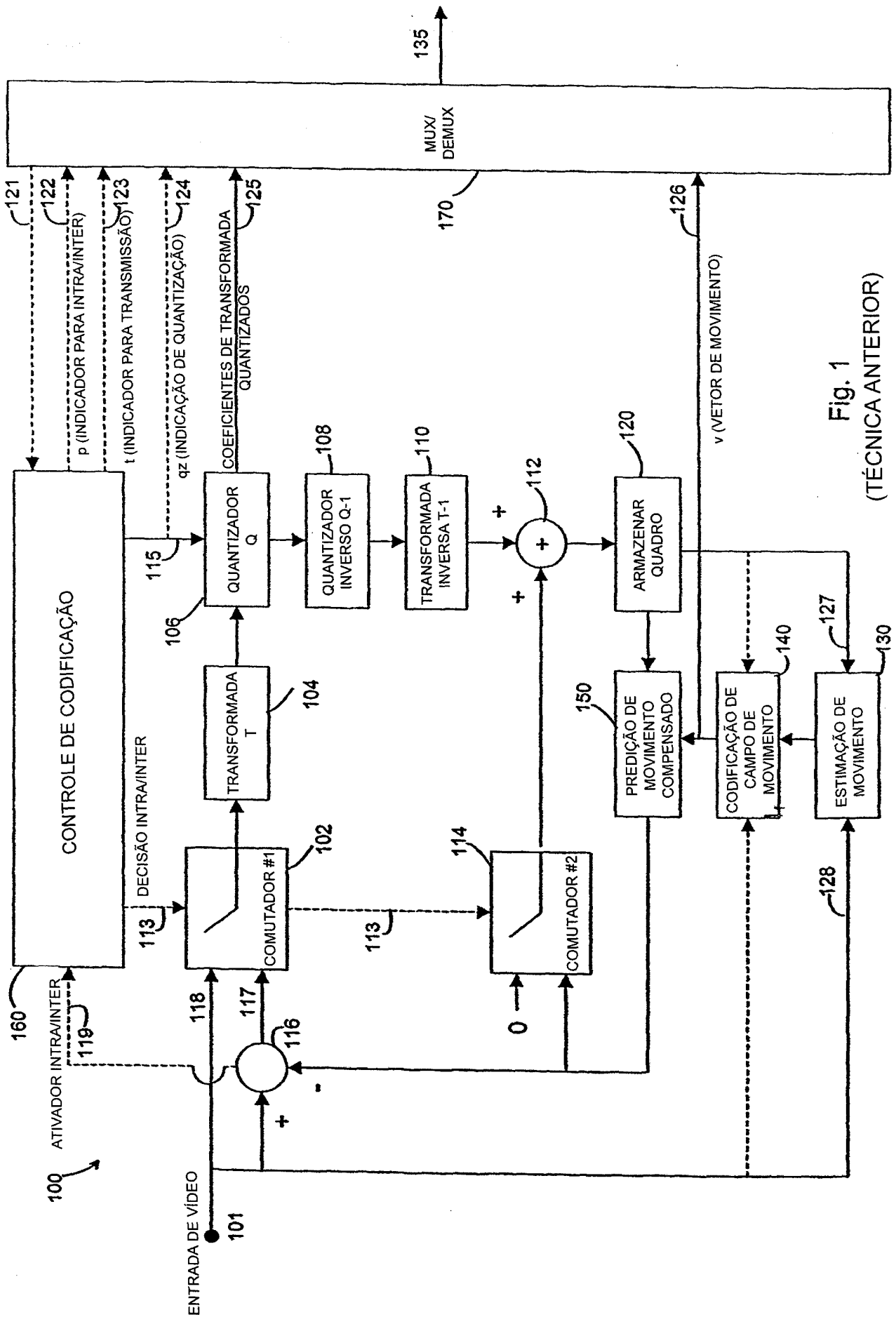


Fig. 1
(TÉCNICA ANTERIOR)

62

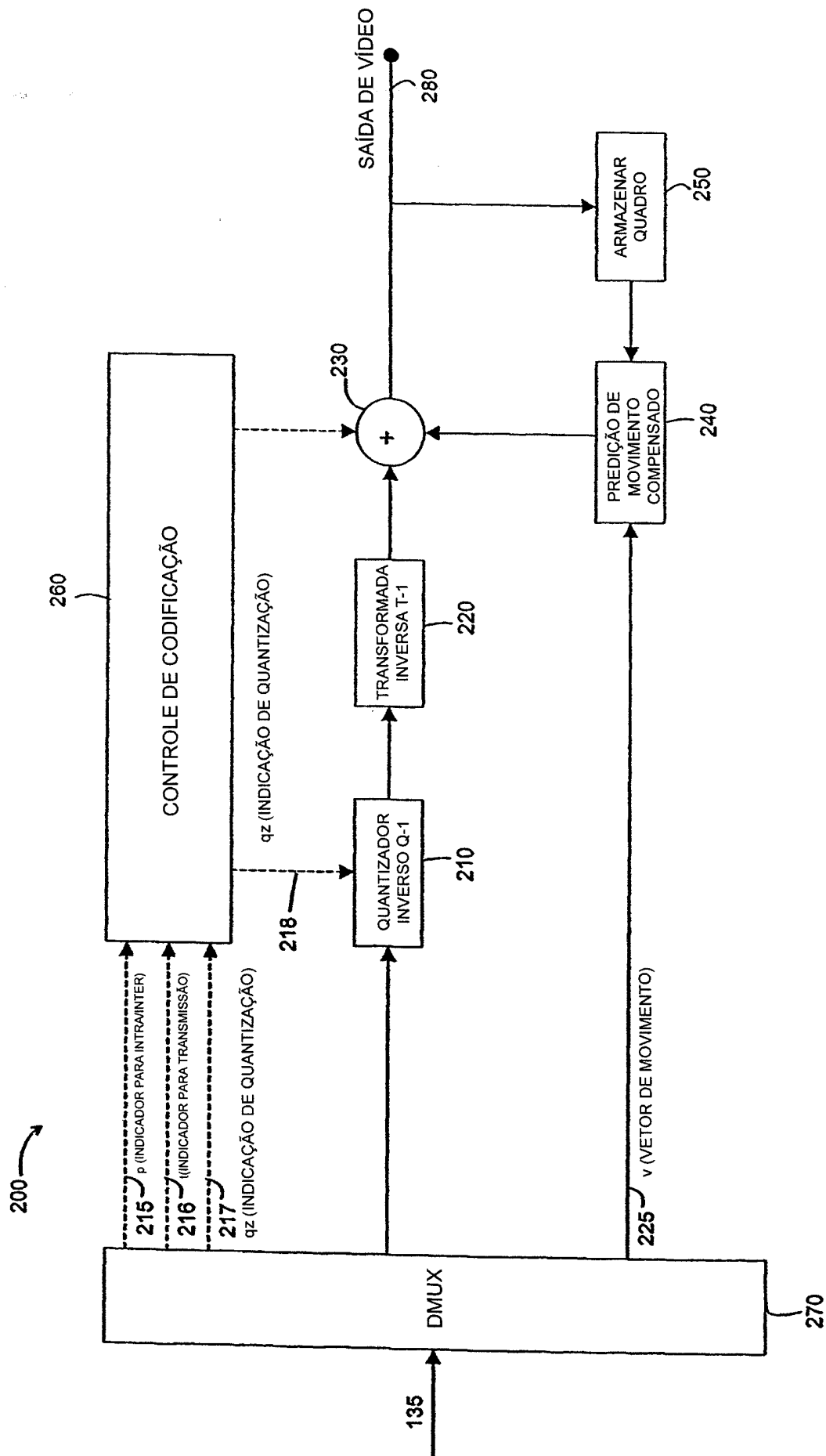
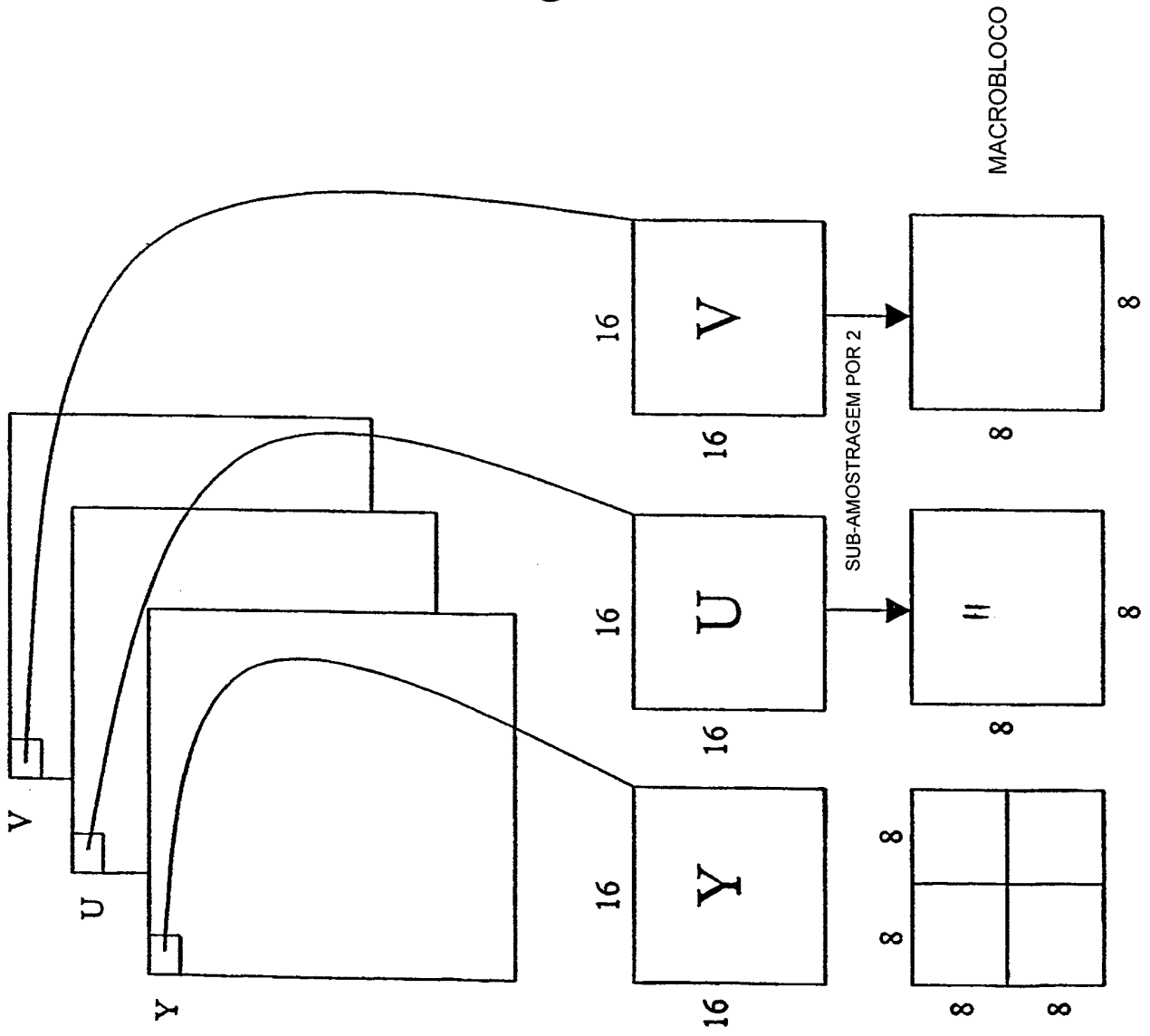


Fig. 2
TÉCNICA ANTERIOR

Fig.3
(TÉCNICA ANTERIOR)



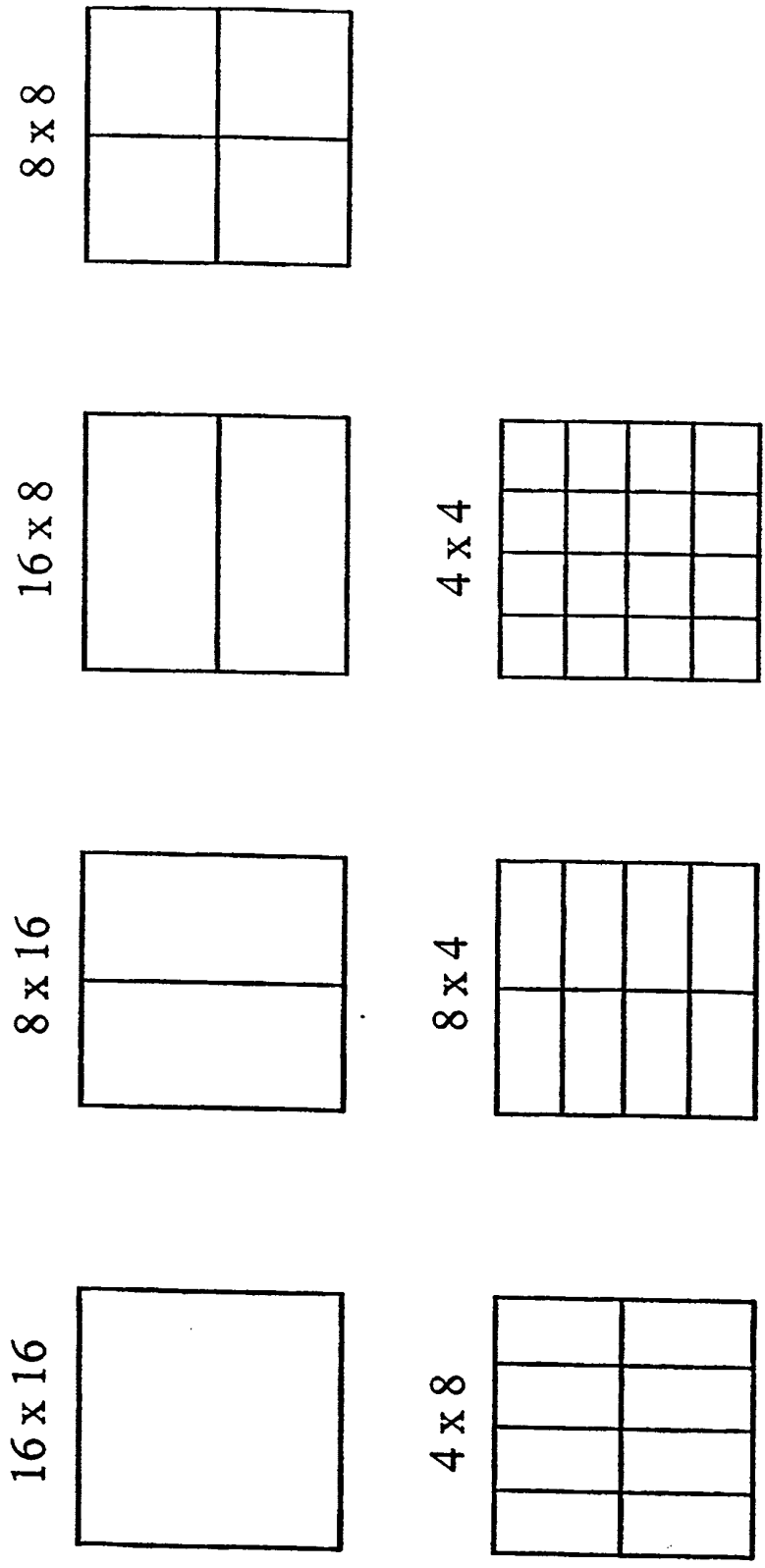


Fig.4

62

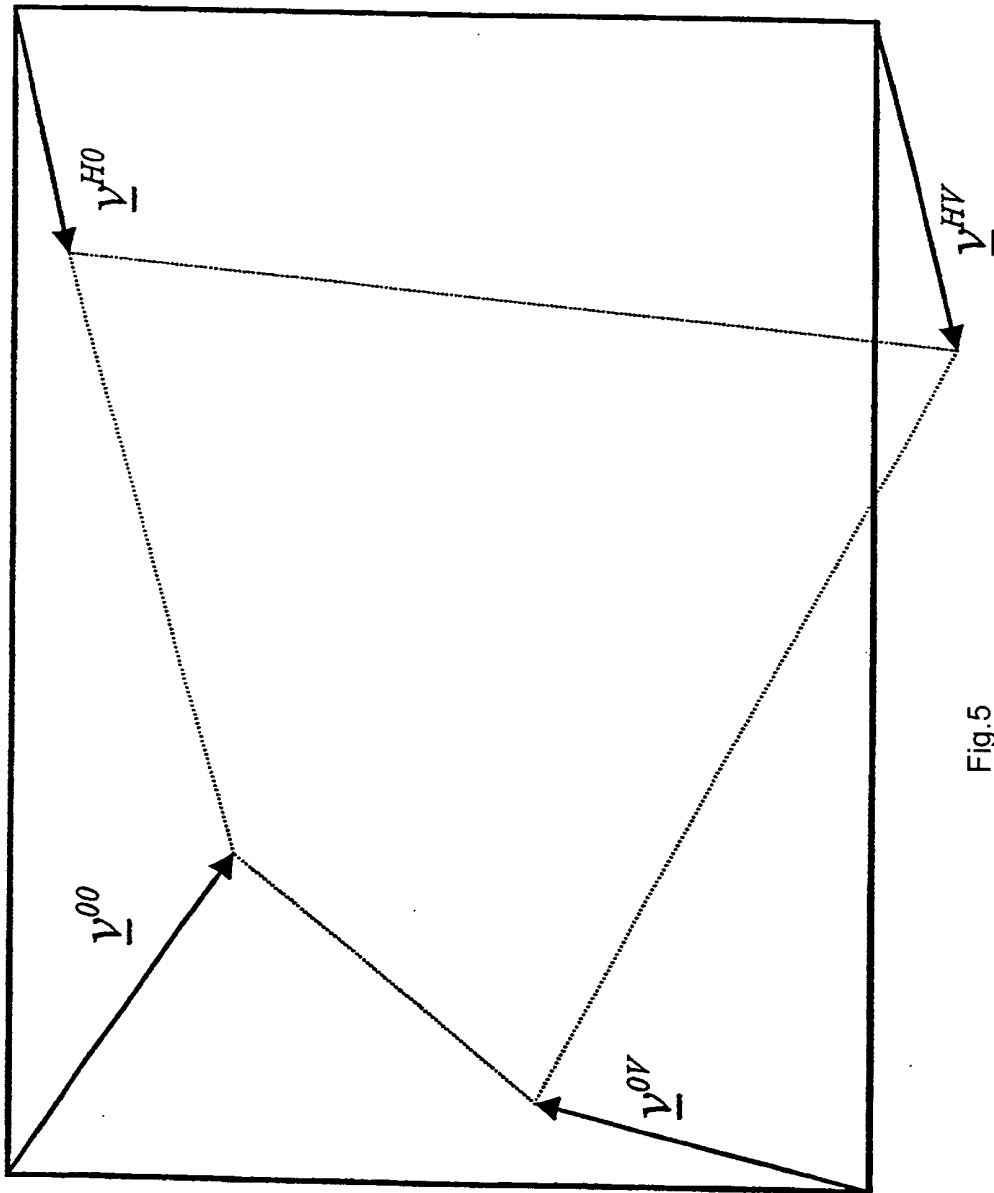


Fig.5

160

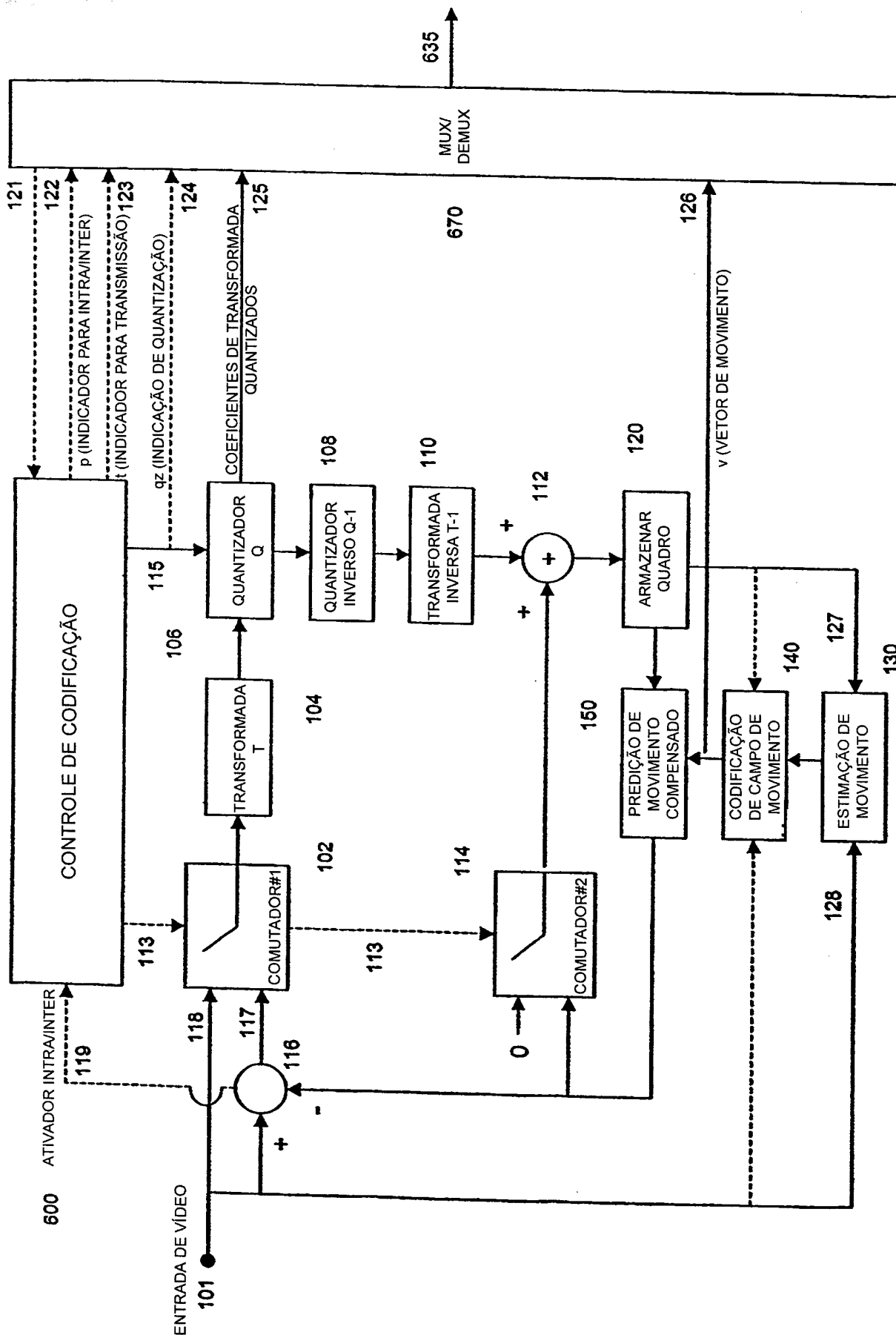


Fig.6

66

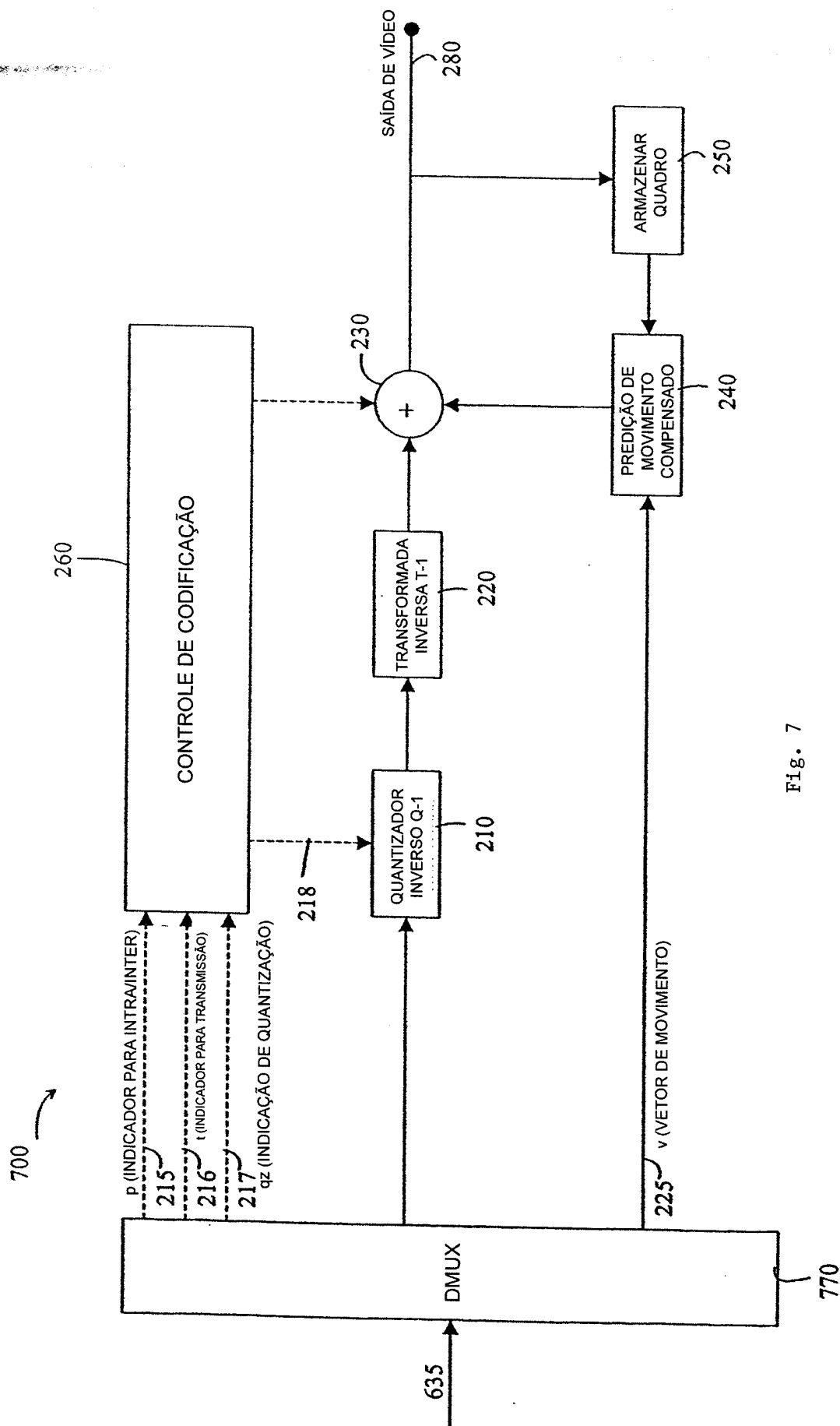


Fig. 7

68
2

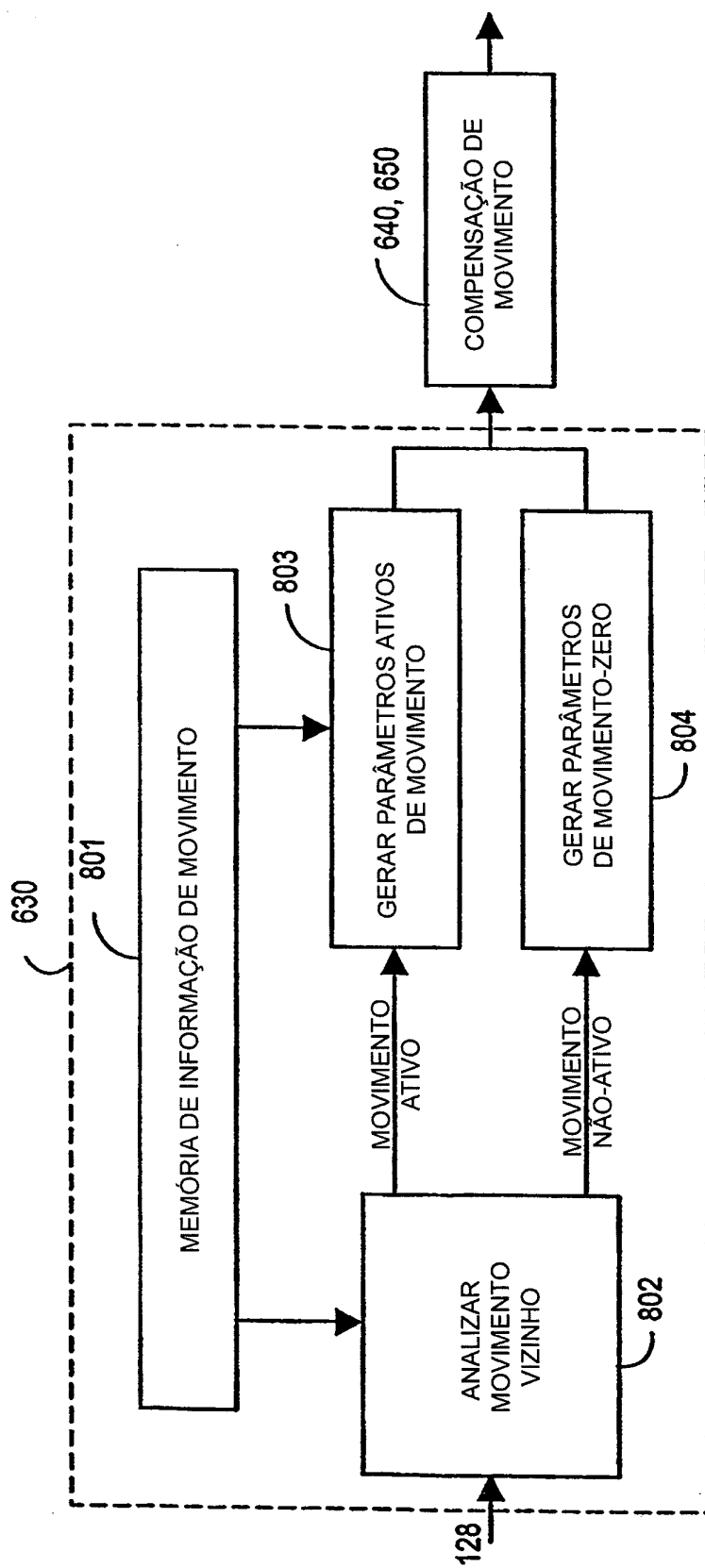


Fig.8

68

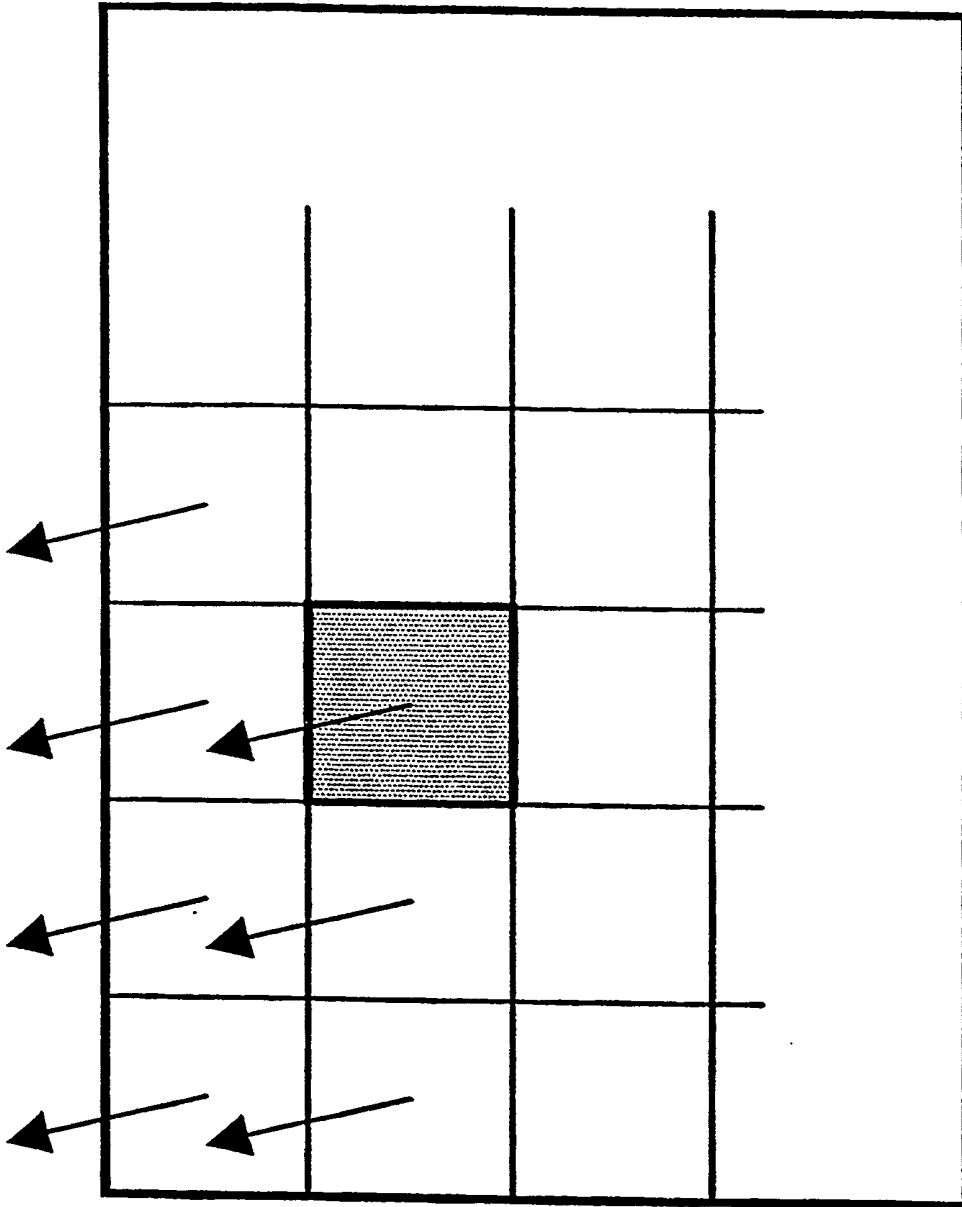


Fig.9

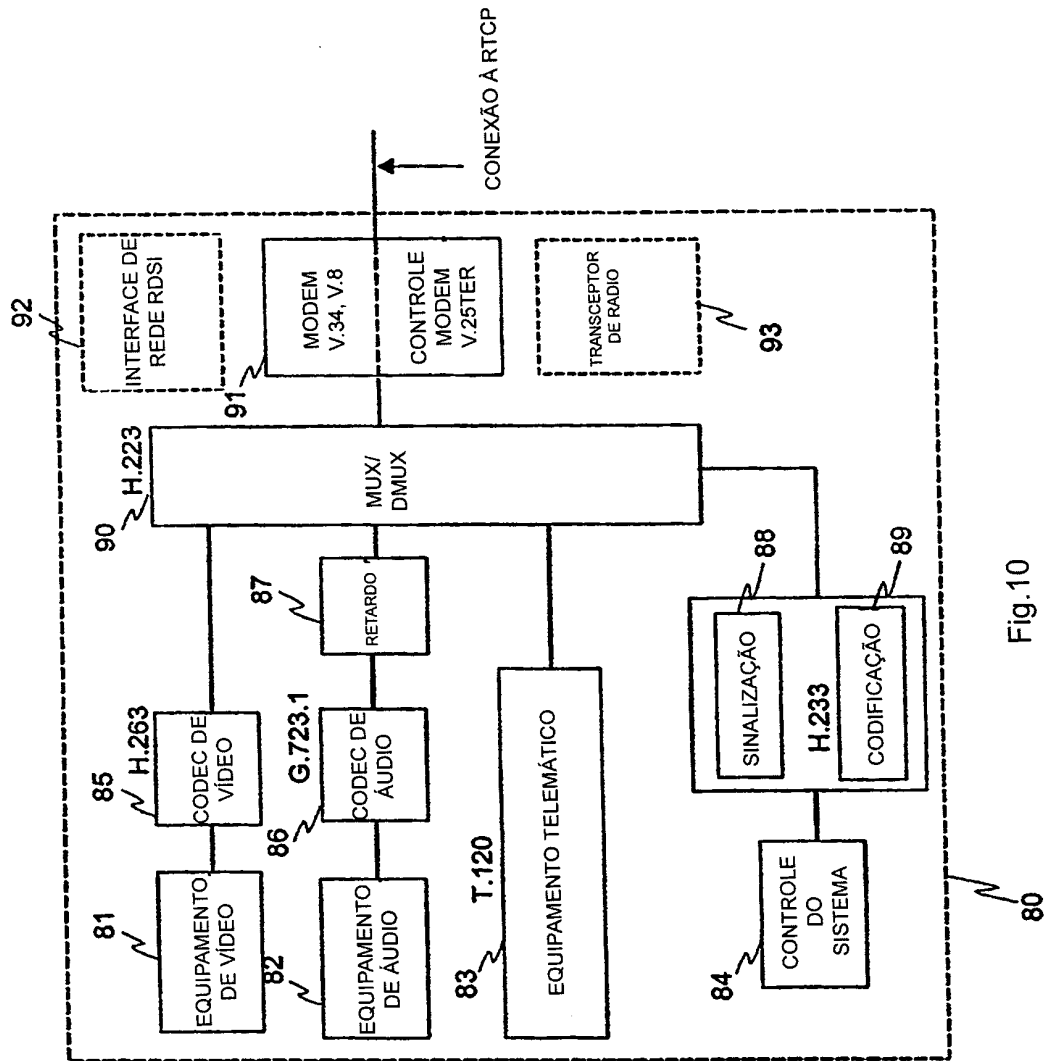


Fig.10