



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

(21) PI 0721077-9 A2



(22) Data de Depósito: 26/12/2007  
(43) Data da Publicação: 01/07/2014  
(RPI 2269)

(51) Int.Cl.:  
H04N 7/32

(54) **Título:** MÉTODO DE CODIFICAÇÃO E MÉTODO DE DECODIFICAÇÃO DE VÍDEO, APARELHOS PARA OS MESMOS, PROGRAMAS PARA OS MESMOS, E MEIO DE ARMAZENAMENTO O QUAL ARMAZENA OS PROGRAMAS

(57) **Resumo:**

(30) **Prioridade Unionista:** 28/12/2006 JP 2006-353628

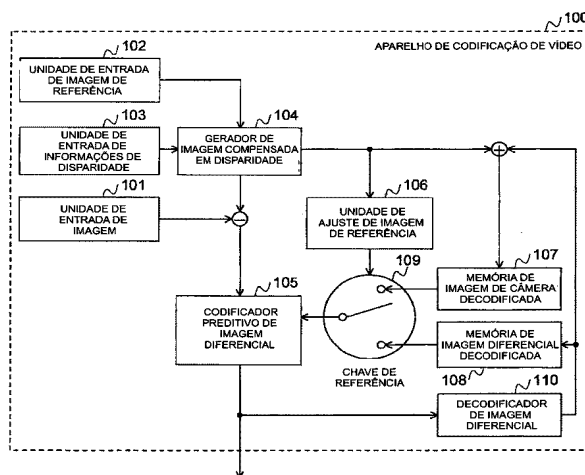
(73) **Titular(es):** Nippon Telegraph and Telephone Corporation

(72) **Inventor(es):** Kazuto Kamikura, Masaki Kitahara, Shinya Shimizu, Yoshiyuki Yashima

(74) **Procurador(es):** Dannemann, Siemsen, Bigler & Ipanema Moreira

(86) **Pedido Internacional:** PCT JP2007074941 de 26/12/2007

(87) **Publicação Internacional:** WO 2008/081810de  
10/07/2008



Relatório Descritivo da Patente de Invenção para **"MÉTODO DE CODIFICAÇÃO E MÉTODO DE DECODIFICAÇÃO DE VÍDEO, APARELHOS PARA OS MESMOS, PROGRAMAS PARA OS MESMOS, E MEIO DE ARMAZENAMENTO O QUAL ARMAZENA OS PROGRAMAS"**.

5 CAMPO DA TÉCNICA

A presente invenção refere-se a técnicas de codificação e decodificação para imagens de vídeo de múltiplos pontos de vista.

Prioridade é reivindicada sobre o Pedido de Patente Japonesa Número 2006-353628, depositado em 28 de Dezembro de 2006, o conteúdo  
10 do qual está aqui incorporado por referência.

ANTECEDENTES DA TÉCNICA

As imagens de vídeo de múltiplos pontos de vista são uma pluralidade de imagens de vídeo obtidas fotografando o mesmo objeto e o seu fundo utilizando uma pluralidade de câmeras. Abaixo, uma imagem de vídeo  
15 obtida por uma única câmera é denominada uma "imagem de vídeo bidimensional", e um conjunto de múltiplas imagens de vídeo bidimensionais obtidas fotografando o mesmo objeto e o seu fundo é denominada uma "imagem de vídeo de múltiplos pontos de vista".

Existe uma forte correlação temporal na imagem de vídeo bidimensional de cada câmera, a qual está incluída em uma imagem de vídeo  
20 de múltiplos pontos de vista. Além disso, quando as câmeras estão sincronizadas umas com as outras, as imagens (feitas pelas câmeras) ao mesmo tempo capturam o objeto e o seu fundo inteiramente no mesmo estado de diferentes posições, de modo que existe uma forte correlação entre as câ-  
25 meras. A eficiência de codificação de codificação de vídeo pode ser aperfeiçoada utilizando esta correlação.

Primeiro, as técnicas convencionais relativas à codificação de imagens de vídeo bidimensionais serão mostradas.

Em muitos métodos conhecidos de codificação de imagens de  
30 vídeo bidimensionais, tais como H.264, MPEG-2, MPEG-4 (os quais são padrões de codificação internacionais), e similares, uma codificação altamente eficiente é executada por meio de compensação de movimento, transforma-

ção ortogonal, quantização, codificação de entropia, ou similares. Por exemplo, em H.264, é possível executar a codificação utilizando uma correlação temporal entre o quadro presente e quadros passados ou futuros.

O Documento Não-patente 1 descreve técnicas detalhadas de compensação de movimento utilizadas em H.264. Suas explicações gerais seguem.

De acordo com a compensação de movimento em H.264, um quadro-alvo de codificação é dividido em blocos de qualquer tamanho, e cada bloco pode ter um vetor de movimento inicial, por meio disto atingindo um alto nível de eficiência de codificação mesmo para uma mudança local em uma imagem de vídeo.

Além disso, como candidatas para uma imagem de referência, quadros passados e futuros (com relação ao quadro presente), os quais já foram codificados, podem ser preparados de modo que cada bloco possa ter um quadro de referência individual, por meio disto implementando um alto nível de eficiência de codificação mesmo para uma imagem de vídeo na qual uma oclusão ocorre devido a uma mudança temporal.

A seguir, um método de codificação de imagens de vídeo de múltiplos pontos de vista convencional será explicado. Como a codificação de imagens de vídeo de múltiplos pontos de vista utiliza uma correlação entre as câmeras, as imagens de vídeo de múltiplos pontos de vista são codificadas altamente eficientemente em um método conhecido o qual utiliza uma "compensação de disparidade" na qual uma compensação de movimento é aplicada a imagens obtidas por diferentes câmeras ao mesmo tempo. Aqui, a disparidade é a diferença entre as posições, nas quais o mesmo ponto sobre um objeto de imagem formada é projetado, sobre os planos de imagem de câmeras as quais estão dispostas em diferentes posições.

A Figura 7 é uma vista esquemática que mostra o conceito de disparidade gerado entre tais câmeras. Isto é, a Figura 7 mostra um estado no qual um observador olha para baixo sobre os planos de imagem de câmeras A e B, cujos eixos geométricos óticos são paralelos um ao outro, de seu lado superior. geralmente, tais pontos, nos quais o mesmo ponto de um

objeto de imagem formada é projetado, sobre os planos de diferentes câmeras, são denominados "pontos correspondentes". Na codificação baseada em compensação de disparidade, com base na relação correspondente acima, cada valor de pixel de um quadro-alvo de codificação é predito utilizando um quadro de referência, e as informações residuais e de disparidade de predição relevantes as quais indicam a relação correspondente são codificadas.

Para cada quadro em uma imagem de vídeo de múltiplos pontos de vista, uma redundância temporal e uma redundância entre as câmeras estão presentes ao mesmo tempo. O Documento Não-patente 2 e o Documento Não-patente 1 (que descreve um aparelho de codificação de imagem de múltiplos pontos de vista) cada um descreve um método para remover ambas as redundâncias simultaneamente.

Nos métodos relevantes, uma predição temporal de uma imagem diferencial entre uma imagem original e uma imagem de disparidade compensada é executada, e um resíduo de compensação de movimento na imagem diferencial é codificado.

De acordo com tais métodos, a redundância temporal, a qual não pode ser removida por uma compensação de disparidade para remover a redundância intercâmeras, pode ser removida utilizando a compensação de movimento. Portanto, uma predição de resíduo, o qual é finalmente codificado, é reduzida, de modo que um alto nível de eficiência de codificação pode ser atingido.

Documento Não-patente 1: ITU-T Rec. H.264/ISO/IEC 11496-10, "Editor's Proposed Draft Text Modifications for Joint Video Specification (ITU-T Rec. H.264/ISO/IEC 11496-10 AVC), Draft 7" Final Committee Draft, Document JVT-022, pp. 10-13, e 62-68, Setembro de 2002.

Documento Não-patente 2: Shinya SHIMIZU, Masaki KITAHARA, Kazuto KAMIKURA e Yoshiyuki YASHIMA, "Multi-view Video Coding based on 3-D Warping with Depth Map", In Proceedings os Picture Coding Symposium 2006, SS3-6, Abril de 2006

Documento de Patente 1: Pedido de Patente Não-Examinada

Japonesa, Primeira Publicação Número H10-191393.

## DESCRIÇÃO DA INVENÇÃO

### PROBLEMA A SER RESOLVIDO PELA INVENÇÃO

De acordo com o método de codificação de vídeo de múltiplos  
5 pontos de vista convencional que aplica uma compensação de movimento a uma imagem diferencial, o resíduo de predição em uma parte que tem tanto a redundância temporal quanto intercâmeras pode ser reduzida, de modo que a codificação possa ser eficientemente executada.

No entanto, se uma oclusão ocorrer entre os objetos de imagem  
10 formada como mostrado na Figura 8, então a presença / ausência de redundância intercâmeras para um único objeto pode mudar dependendo do tempo.

A Figura 8 mostra um exemplo de oclusão interobjetos em imagens as quais são obtidas pelas câmeras A e B em cada um dos tempos a e  
15 b.

A Figura 9 mostra um exemplo da imagem de resíduo de compensação de disparidade quando existe uma oclusão.

De acordo com o método descrito no Documento Não-patente 2, para cada pixel de um quadro de referência utilizado em compensação de  
20 disparidade, as informações de disparidade as quais indicam um ponto correspondente em outro quadro estão providas. Portanto, se a imagem da câmera B for compensada em disparidade da imagem da câmera A (ver Figura 8), uma imagem diferencial como mostrado na Figura 9 é obtida.

Na Figura 9, a profundidade de cor indica o valor do sinal diferencial, de modo que quanto mais próximo do branco, menor a diferença  
25 relevante.

Aqui, as informações de disparidade, as quais indicam a qual parte na imagem da câmera B cada pixel no quadro de referência (isto é, a imagem da câmera A) corresponde, são utilizadas. Portanto, nenhuma com-  
30 pensação de disparidade é executada em cada parte a qual está presente a imagem da câmera B mas não está presente na imagem da câmera A. Consequentemente, a cor de cada parte (na imagem relevante) para a qual um

ponto correspondente pode ser obtido torna-se próxima de branco. Em contraste, na parte R (dentro de um círculo tracejado na Figura 9) para a qual nenhum ponto correspondente pode ser obtido, a imagem original permanece não-mudada como o sinal diferencial.

5                   No entanto, devido a uma diferença em sensibilidade entre as câmeras ou uma influência de reflexão, o sinal diferencial não torna-se completamente zero mesmo em cada parte para a qual um ponto correspondente pode ser obtido. Em uma área tal como R na Figura 9, é óbvio que o resíduo de predição não pode ser reduzido mesmo por compensação de movimento executada de uma imagem diferencial em outro tempo tal como o  
10                   tempo a na Figura 9.

                  Em um método simples para resolver o problema acima, cada bloco refere a um quadro diferente. Isto é, como o quadro de referência, não somente a imagem decodificada da imagem diferencial, mas também a ima-  
15                   gem decodificada de uma imagem de câmera final obtida pela adição de uma imagem compensada em disparidade à imagem diferencial pode ser utilizada, e uma das duas imagens decodificadas pode ser intercambiavelmente referida em cada bloco.

                  Consequentemente, mesmo na área R da Figura 9 em que a  
20                   imagem de câmera original permanece, o resíduo de predição pode ser reduzido pela execução de compensação de movimento da imagem decodificada da imagem de câmera.

                  No entanto, em tal método, é necessário codificar as informações para indicar o quadro de referência de cada bloco, o que causa um  
25                   aumento na quantidade de código.

                  Além disso, o método acima é ineficaz se uma parte que tem uma oclusão e uma parte que não tem oclusão estiverem simultaneamente presentes em um bloco.

                  Por outro lado, o método descrito no Documento Não-patente 1,  
30                   quando codificando um quadro-alvo de codificação, ao invés de codificar o quadro inteiro sujeitando a imagem diferencial à compensação de movimento, é possível seleccionar para cada bloco se (i) a imagem diferencial é sujeita

à compensação de movimento, (ii) somente uma compensação de disparidade é executada, ou (iii) somente uma compensação de movimento é executada.

Consequentemente, mesmo uma parte onde uma oclusão ocorre entre as câmeras, o resíduo de predição pode ser eficientemente reduzido se existir uma redundância temporal.

No entanto, no método acima, é necessário codificar as informações para indicar qual método de predição foi utilizado em cada bloco, isto é, codificar uma grande quantidade de informações adicionais apesar do resíduo de predição poder ser reduzido. Portanto, é impossível atingir um alto nível de eficiência de codificação.

Mais ainda, como somente um método de predição pode ser selecionado em cada bloco, o método relevante é ineficaz se a presença / ausência de oclusão não for fixada em cada parte de um bloco.

À luz das circunstâncias acima, um objeto da presente invenção é prover novas técnicas de codificação e de decodificação de vídeo para atingir um alto nível de eficiência de codificação pela implementação de uma predição apropriada sem aumentar a quantidade de código requerida para a predição.

## 20 MEIOS PARA RESOLVER O PROBLEMA

De modo a resolver os problemas acima, a presente invenção provê um método de codificação de vídeo para gerar, com base nas informações de disparidade entre uma imagem de câmera de referência já codificada e uma imagem de câmera-alvo de codificação que corresponde à imagem de câmera de referência, uma imagem compensada em disparidade por meio de predição de imagem intercâmeras; e codificar uma imagem diferencial entre a imagem de câmera-alvo de codificação e a imagem compensada em disparidade. O método tem uma etapa de selecionar, para cada unidade de seção predeterminada (por exemplo, cada pixel) sobre a imagem diferencial, um dos seguintes dois grupos de imagem como um alvo de referência, determinando se existe ou não uma imagem compensada em disparidade na posição correspondente, em outras palavras se o pixel correspondente na

imagem compensada em disparidade tem ou não um valor efetivo:

(i) um grupo de imagens diferenciais decodificadas, o qual é obtido pela decodificação de uma imagem diferencial entre a imagem de câmera já codificada e a imagem compensada em disparidade (isto é, um conjunto de imagens diferenciais decodificadas o qual inclui uma imagem diferencial decodificada a qual foi obtida em outro momento e já foi codificada, ou uma imagem diferencial decodificada de uma parte já codificada da imagem diferencial para a imagem de câmera-alvo de codificação); e

(ii) um grupo de imagens de câmera decodificadas, o qual é obtido pela decodificação da imagem de câmera já codificada (um conjunto de imagens de câmera decodificadas as quais são cada uma representada pela soma de cada imagem diferencial decodificada (a qual pode ser uma imagem diferencial decodificada em outro momento) e uma imagem compensada em disparidade correspondente).

De acordo com o método acima, uma codificação preditiva que utiliza uma imagem de referência apropriada para cada unidade de seção desejada pode ser executada sem acrescentar novas informações adicionais sobre a seleção para o modo ou a imagem de referência. Isto é, é possível reduzir o elemento residual a ser codificado, sem aumentar a quantidade de código requerida para as informações adicionais, por meio disto implementando um alto nível de eficiência de codificação.

No método acima, as informações de disparidade podem ser providas como um vetor de disparidade, ou representadas como informações geométricas tais como uma distância da câmera para o objeto de imagem formada, isto é, estas podem ser providas em qualquer formato pelo qual as relações correspondentes entre as imagens relevantes podem ser obtidas.

Quando codificando a imagem diferencial utilizando um alvo de referência designado, a codificação pode ser executada utilizando um método de predição de sinal em um bloco de processamento de codificação.

Na codificação preditiva geral, diferentes alvos de referência provavelmente têm diferentes métodos de predição ótimos. Portanto, se um

bloco de processamento de codificação incluir pixels ou áreas as quais têm diferentes alvos de referência, o método de predição é selecionado para cada alvo de referência, e informações adicionais as quais indicam uma pluralidade de métodos de predição são codificadas.

- 5                   No entanto, se a codificação for executada utilizando um único método de predição em um bloco de processamento de codificação, como acima descrito, as informações as quais indicam somente um método de predição deveriam ser codificadas para cada bloco de processamento de codificação, por meio disto reduzindo a quantidade de código requerida para a codificação inteira. Neste caso, para uma degradação prevista na eficiência de predição, a eficiência de predição não deve ser consideravelmente degradada em consideração de um caso como mostrado na Figura 9, isto é, quando executando uma predição de vídeo pela aplicação de um único método (isto é, uma compensação de movimento que utiliza um único vetor de movimento) mesmo em uma parte de limite.

- 15                   Além disso, em consideração do equilíbrio entre a eficiência de predição e a quantidade de código requerida para as informações adicionais, é possível selecionar se um único método de predição é utilizado em cada bloco, ou um método de predição individual é atribuído a cada alvo de referência. Quando considerando tal equilíbrio entre a eficiência de predição e a quantidade de código requerida para as informações adicionais, é possível implementar uma codificação mais flexível que tem um nível mais alto de eficiência de codificação.

- 20                   Neste caso, o lado de decodificação pode determinar dos dados codificados se o método de predição foi mudado ou não, de modo que o lado de decodificação possa decodificar apropriadamente a imagem relevante.

#### EFEITO DA INVENÇÃO

- 25                   De acordo com a presente invenção, quando sujeitando uma imagem diferencial obtida por compensação de disparidade a uma codificação preditiva, nenhuma informação nova adicional é utilizada, e uma imagem de referência apropriada é utilizada através de uma operação de troca para cada pixel. Portanto, o resíduo para o alvo de codificação pode ser reduzido

sem aumentar a quantidade de código requerida para as informações adicionais, por meio disto implementando uma codificação de vídeo eficiente da imagem de vídeo de múltiplos pontos de vista.

### BREVE DESCRIÇÃO DOS DESENHOS

5                    Figura 1 é um diagrama que mostra um aparelho de codificação de vídeo como uma modalidade da presente invenção.

Figura 2 é um fluxograma de codificação de vídeo executado na modalidade.

10                    Figura 3 é um fluxograma detalhado para a determinação de modo de predição na modalidade.

Figura 4 é um diagrama que mostra um aparelho de decodificação de vídeo como uma modalidade da presente invenção.

Figura 5 é um fluxograma de decodificação de vídeo executado na modalidade.

15                    Figura 6 é um fluxograma detalhado da decodificação de dados codificados para a imagem diferencial na modalidade.

Figura 7 é um diagrama esquemático que mostra o conceito de disparidade gerado entre as câmeras.

20                    Figura 8 é um diagrama que mostra um exemplo no qual uma oclusão ocorre entre objetos de imagem formada.

Figura 9 é um diagrama que mostra um exemplo da imagem residual de compensação de disparidade quando existe uma oclusão.

### LISTAGEM DE REFERÊNCIA

|        |                                                  |
|--------|--------------------------------------------------|
| 100    | aparelho de codificação de vídeo                 |
| 25 101 | unidade de entrada de imagem                     |
| 102    | unidade de entrada de imagem de referência       |
| 103    | unidade de entrada de informações de disparidade |
| 104    | gerador de imagem compensada em disparidade      |
| 105    | codificador preditivo de imagem diferencial      |
| 30 106 | unidade de ajuste de imagem de referência        |
| 107    | memória de imagem de câmera decodificada         |
| 108    | memória de imagem diferencial decodificada       |

|    |     |                                                  |
|----|-----|--------------------------------------------------|
|    | 109 | chave de referência                              |
|    | 110 | decodificador de imagem diferencial              |
|    | 200 | aparelho de decodificação de vídeo               |
|    | 201 | unidade de entrada de dados codificados          |
| 5  | 202 | unidade de entrada de imagem de referência       |
|    | 203 | unidade de entrada de informações de disparidade |
|    | 204 | gerador de imagem compensada em disparidade      |
|    | 205 | unidade de ajuste de imagem de referência        |
|    | 206 | memória de imagem de câmera decodificada         |
| 10 | 207 | memória de imagem diferencial decodificada       |
|    | 208 | chave de referência                              |
|    | 209 | decodificador de imagem diferencial              |

#### MELHOR MODO PARA EXECUTAR A INVENÇÃO

Antes da explicação de modalidades da presente invenção, um  
 15 sumário da presente invenção será descrito.

Em técnicas convencionais para a codificação de múltiplos pontos de vista, é necessário codificar e transmitir as informações as quais indicam a imagem selecionada. Em contraste, a presente invenção está caracterizada pelo fato de que as informações as quais indicam a imagem selecionada podem ser decodificadas utilizando outras informações. Consequentemente, quando a presente invenção tem a mesma quantidade de código que as técnicas convencionais, a presente invenção tem uma melhor qualidade de imagem em comparação com as técnicas convencionais.

De modo a codificar uma imagem de vídeo de múltiplos pontos de vista, se uma diferença entre uma imagem de vídeo obtida por uma câmera e uma imagem correspondente gerada por meio de compensação de disparidade de outra câmera for obtida, e a codificação é executada em consideração de correlação temporal com relação à imagem de vídeo diferencial, então nas técnicas convencionais, uma parte de oclusão tal como R na  
 25  
 30 Figura 9 não pode ser eficientemente codificada.

Em contraste, na presente invenção, uma parte de oclusão tal como R na Figura 9 pode ser eficientemente codificada sujeitando a parte

relevante à codificação preditiva a qual utiliza, não uma imagem de vídeo diferencial de uma imagem gerada por compensação de disparidade, mas uma imagem decodificada da imagem de vídeo original.

Além disso, em um método facilmente previsto para resolver um problema tal que a parte de oclusão não pode ser eficientemente codificada, uma técnica de referência de múltiplos quadros empregada em H.264 é utilizada de modo que uma codificação preditiva é executada atribuindo uma imagem de referência individual a cada bloco.

No entanto, em tal caso, (i) as informações para designar uma imagem de referência para cada bloco são codificadas, o que causa um aumento na quantidade de código, e (ii) a troca da imagem de referência pode ser executada somente por uma unidade de bloco. Portanto, o método relevante é ineficaz quando um bloco tem tanto uma parte de oclusão quanto uma parte de não oclusão.

Em contraste, na presente invenção, a troca da imagem de referência é executada utilizando as informações obtidas quando gerando uma imagem por compensação de disparidade. Portanto, é desnecessário codificar novas informações. Mais ainda, de acordo com as informações utilizadas, a presença / ausência de oclusão pode ser determinada para cada pixel. Portanto, a troca da imagem de referência pode ser executada por uma unidade de pixel.

De modo a implementar as funções acima, na presente invenção, as informações Ref as quais indicam se uma imagem compensada em disparidade Syn de um alvo de codificação pode ou não ser gerada utilizando uma imagem obtida por outra câmera são geradas, e a troca da imagem de referência é executada utilizando as informações Ref.

Um resultado de determinação de que uma imagem compensada em disparidade Syn pode ser gerada significa que o mesmo objeto aparece em uma imagem de outra câmera, e assim nenhuma oclusão ocorreu.

Em contraste, um resultado de determinação de que uma imagem compensada em disparidade Syn não pode ser gerada significa que o mesmo objeto não aparece em uma imagem de outra câmera, e assim uma

oclusão ocorreu.

Isto é, a troca da imagem de referência é executada utilizando as informações as quais indicam se a imagem compensada em disparidade Syn pode ser gerada ou não, de modo que é possível resolver o problema  
5 que nenhuma predição apropriada pode ser executada em uma parte de oclusão tal como R na Figura 9.

A técnica de referência de múltiplos quadros utilizada em H.264 é também conhecida como um método para selecionar uma imagem de referência utilizada para a codificação dentre um pluralidade de candidatas de  
10 imagem de referência. Neste método conhecido, é necessário codificar e transmitir as informações que indicam qual imagem foi selecionada.

Em contraste, na presente invenção, as informações utilizadas para selecionar uma imagem de referência são geradas utilizando as informações as quais são utilizadas para gerar uma imagem compensada em  
15 disparidade e a qual deve ser transmitida também quando utilizando o método conhecido acima. Portanto, é desnecessário transmitir as informações adicionais utilizadas para trocar a imagem de referência, por meio disto reduzindo a quantidade de código.

Nas modalidades posteriormente mostradas, as informações as  
20 quais são utilizadas para selecionar uma imagem de referência e são geradas com base em informações utilizadas para gerar uma imagem compensada em disparidade são representadas como as informações Ref, e as mesmas informações podem ser utilizadas nos lados de codificação e de decodificação (por exemplo, as informações no lado de codificação são geradas na etapa A6 na Figura 2, e as informações no lado de decodificação  
25 são geradas na etapa C4 na Figura 5).

Além disso, se a imagem de referência for selecionada para cada pixel em um método conhecido tal como a técnica de referência de múltiplos quadros em H.264 ou similares, a quantidade de código requerida para  
30 um sinal de erro de predição a ser codificado pode ser reduzida. No entanto, em tal caso, é necessário codificar as informações utilizadas para trocar a imagem de referência para cada pixel, o que causa um aumento na quanti-

dade de código. Isto é, neste caso, a imagem de referência pode ser trocada somente por uma unidade de bloco de modo a reduzir a quantidade total de código.

Em contraste, na presente invenção, as informações para trocar a imagem de referência não são umas a serem adicionalmente transmitidas como acima descrito. Portanto, quando gerando tais informações em correspondência com cada pixel, a troca da imagem de referência para cada pixel pode ser executada sem aumentar a quantidade de código; por meio disto reduzindo a quantidade de código requerida para o sinal de erro de predição.

Abaixo, a presente invenção será explicada em detalhes de acordo com as modalidades.

Nas modalidades, é assumido que uma imagem de vídeo de múltiplos pontos de vista obtida utilizando duas câmeras é codificada, em que a imagem da câmera B é codificada utilizando a imagem da câmera A como a imagem de referência.

A Figura 1 mostra um aparelho de codificação de vídeo como uma modalidade da presente invenção.

O aparelho de codificação de vídeo 100 inclui uma unidade de entrada de imagem 101 para inserir uma imagem (como uma imagem-alvo de codificação) obtida pela câmera B no aparelho; uma unidade de entrada de imagem de referência 102 para inserir uma imagem decodificada (como uma imagem de referência) da câmera A no aparelho; uma unidade de entrada de informações de disparidade 103 para inserir as informações de disparidade, as quais indicam para qual posição sobre a imagem-alvo de codificação cada pixel na imagem de referência corresponde, no aparelho; um gerador de imagem compensada em disparidade 104 para gerar uma imagem compensada em disparidade na posição da câmera B pela utilização da imagem de referência e das informações de disparidade; um codificador preditivo de imagem diferencial 105 para sujeitar uma imagem diferencial entre a imagem-alvo de codificação e a imagem compensada em disparidade para uma codificação preditiva; uma unidade de ajuste de imagem de referência 106 para atribuir uma grupo de imagens de referência a cada pixel

com base no estado da imagem compensada em disparidade; uma memória de imagem de câmera decodificada 107 para armazenar as imagens decodificadas (como um grupo de imagens de referência) de imagens obtidas pela câmera B; uma memória de imagem diferencial decodificada 108 para armazenar as imagens decodificadas (como outro grupo de imagens de referência) das imagens diferenciais codificadas; uma chave de referência 109 para trocar a imagem de memória de referência com base nas informações de ajuste sobre o grupo de imagens de referência; e um decodificador de imagem diferencial 110 para decodificar cada imagem diferencial codificada.

As Figuras 2 e 3 são fluxogramas executados pelo aparelho de codificação de vídeo 100 que tem a estrutura acima. Referindo aos fluxogramas, a operação executada pelo aparelho de codificação de vídeo 100 serão explicadas em detalhes.

Como mostrado no fluxo de operação da Figura 2 a qual mostra uma operação de codificação geral, uma imagem da câmera B é inserida pela unidade de entrada de imagem 101 no aparelho de codificação de vídeo 100 (ver etapa A1).

Aqui, uma imagem decodificada da câmera A, a qual tem o mesmo tempo de exibição (isto é, o mesmo tempo de formação de imagem) que a imagem de entrada da câmera B, é inserida pela unidade de entrada de imagem de referência 102, e as informações de disparidade entre as imagens no tempo relevante são inseridas pela unidade de entrada de informações de disparidade 103.

Abaixo, a imagem de entrada da câmera B é denominada uma "imagem de câmera-alvo de codificação", e a imagem decodificada da câmera A é denominada uma "imagem de câmera de referência".

Com base na entrada de imagem de câmera de referência pela unidade de entrada de imagem de referência 102 e nas informações de disparidade inseridas pela unidade de entrada de informações de disparidade 103, uma imagem compensada em disparidade é gerada pelo gerador de imagem compensada em disparidade 104 (ver etapa A2).

A geração de imagem compensada em disparidade é executada

em um modo tal que o valor de pixel de cada pixel sobre a imagem de câmera de referência é utilizado como o valor de pixel de um pixel correspondente sobre a imagem compensada em disparidade.

Neste processo, se uma pluralidade de pontos sobre a imagem de câmera de referência corresponder ao mesmo ponto sobre a imagem compensada em disparidade, então relações de frente para trás entre os objetos de imagem formada são determinadas, e o valor de pixel de um ponto sobre um objeto de imagem formada mais próximo da câmera é utilizado. Por exemplo, se as câmeras têm eixos geométricos óticos paralelos, é possível determinar que quanto maior a disparidade, mais próximo o ponto relevante está da câmera.

Além disso, como as informações de disparidade indicam um ponto correspondente sobre a imagem de câmera-alvo de codificação para cada pixel sobre a imagem de câmera de referência, não existe nenhuma imagem compensada em disparidade para uma parte a qual está ausente na imagem de câmera de referência.

Pela utilização da imagem compensada em disparidade acima obtida, a imagem de câmera-alvo de codificação é codificada para cada bloco de processamento de codificação (ver etapas A3 até A13).

No fluxo corrente, "blk" é um índice para cada bloco, e maxBlk indica o número total de blocos em uma imagem. Isto é, após blk ser inicializado para zero (ver etapa A3), o processo seguinte (etapas A4 até A11) é repetidamente executado enquanto blk é incrementado por 1 (ver etapa A12), até blk atingir o número total de blocos maxBlk (ver etapa A13).

Primeiro, é definido para o bloco blk que Org indica a imagem de câmera-alvo de codificação e Syn indica a imagem compensada em disparidade (ver etapa A4). A diferença Org-Syn é computada para cada pixel que pertence ao bloco blk, de modo que uma imagem diferencial Sub é gerada (ver etapa A5).

Para cada parte de pixel que não tem uma imagem compensada em disparidade, o valor de pixel de Syn é ajustado para 0 quando computando a imagem diferencial.

A seguir, a unidade de ajuste de imagem de referência 106 ajusta as informações Ref, as quais indicam uma referência utilizada quando sujeitando a imagem diferencial de cada pixel no bloco blk a uma codificação preditiva como abaixo mostrado (ver etapa A6).

5  $\forall \text{pix} \in \{0, 1, \dots, \text{maxPix}-1\}$

Ref[pix]=0 Quando nenhuma imagem compensada em disparidade está armazenada em Syn[pix]

Ref[pix]=1 Quando uma imagem compensada em disparidade está armazenada em Syn[pix]

10 em que pix é um índice que indica cada pixel no bloco blk, maxPix indica o número total de pixels no bloco blk. Além disso, as informações de unidade de bloco ao qual "[pix]" está anexado indicam as informações do pixel na posição de índice de pixel pix.

Além disso, apesar de qualquer método poder ser utilizado para  
15 determinar se uma imagem compensada em disparidade é ajustada como Syn de modo a gerar Ref, um seu exemplo será abaixo mostrado.

Antes de gerar a imagem compensada em disparidade Syn, o valor de pixel de cada pixel é inicializado a um valor absolutamente inatingível (por exemplo, -1). A imagem compensada em disparidade Syn é gerada  
20 pela atualização do valor de pixel em um modo tal que o valor de pixel de cada pixel na imagem de câmera de referência é determinado como o valor de pixel na posição correspondente sobre a imagem compensada em disparidade. Ref é então gerado em um modo tal que se Syn atribuído a cada pixel for -1, Ref é ajustado para 0, e se Syn tiver um valor outro que -1, Ref é  
25 ajustado para 1. Consequentemente, o Ref alvo pode ser gerado em uma correspondência para Syn de um para um.

No método acima na computação da imagem diferencial Sub na etapa A5 acima descrita, em cada parte em que Syn é -1, o valor de pixel é considerado como 0 e Sub é ajustado para o valor de Org.

30 Em outro método, antes de computar a imagem diferencial Sub, Ref é gerado de acordo com o método acima descrito ou similar; em cada parte em que Syn é -1, Syn é reescrito como 0; e então a imagem diferencial

Sub é computada computando Org-Syn.

Aqui, para conveniência de explicações, o Syn total, o qual também inclui cada parte em que nenhuma imagem compensada em disparidade está presente e o valor de pixel é considerado como 0, é denominado a  
5 imagem compensada em disparidade.

A seguir, um modo de predição, o qual provê o custo de distorção de taxa mínimo quando sujeitando a imagem diferencial Sub à codificação preditiva pela utilização do Ref acima, é obtido e ajustado como PMODE (ver etapa A7).

10 Quando "sad" é a soma de valores absolutos de erros de predição em um modo de predição, e "bin" é um valor predito da quantidade de código requerida para codificar as informações as quais indicam o modo de predição, o custo de distorção de taxa denominado "cost" é computado pela seguinte fórmula, em que  $\lambda$  é um multiplicador de Lagrange não identificado  
15 e é um valor predeterminado.

$$\text{cost} = \text{sad} + \lambda \cdot \text{bin}$$

Pela utilização do PMODE acima determinado, a imagem diferencial Sub do bloco blk é realmente sujeita à codificação preditiva (ver etapa A8)

20 A determinação de modo de predição e a codificação real são executadas no codificador preditivo de imagem diferencial 105. O resultado de codificação inclui, não somente os dados codificados da imagem relevante, mas também os dados codificados de informações (por exemplo, PMODE) requeridos para a codificação.

25 O resultado de codificação é emitido do aparelho de codificação de vídeo 100, e também decodificado no decodificador de imagem diferencial 110 (ver etapa A9).

Aqui, Dec indica a imagem decodificada da imagem diferencial Sub do bloco blk, e é armazenado na memória de imagem diferencial decodificada 108 de modo a utilizar Dec na predição intraquadro executada  
30 quando codificando outro bloco ou na predição interquadros executada quando codificando um quadro em outro tempo (ver etapa A10).

Além disso, a imagem de câmera decodificada do bloco blk é gerada pela computação da soma de Dec e Syn, e é armazenada na memória de imagem de câmera decodificada 107 (ver etapa A11).

A operação acima descrita é aplicada a cada bloco.

5 A Figura 3 mostra um fluxo de operação detalhado do processo de determinação de modo de predição (etapa A7) executado no codificador preditivo de imagem diferencial 105.

No fluxo de operação, um custo de distorção de taxa denominado "cost" indicado pela fórmula seguinte é computado para cada modo de predição, e um modo de predição o qual provê o valor mínimo é determinado.

$$\text{cost} = \text{SAD} + \lambda \cdot \text{code}(\text{pmode})$$

em que SAD é a soma de diferenças absolutas para cada pixel entre a imagem predita e a imagem original; pmode indica um índice de modo de predição; e  $\text{code}(\alpha)$  é uma função a qual retorna um valor predito da quantidade de código requerida para representar as informações sobre o modo de predição provido  $\alpha$ .

Como mostrado no fluxo de operação, após o índice de modo de predição pmode ser inicializado para 0 (ver etapa B1), o seguinte processo (etapas B2 até B16) é repetidamente executado enquanto pmode é incrementado por 1 (ver etapa B17), até pmode atingir o número "maxPmode" de modos de predição (ver etapa B18), por meio disto determinando um modo de predição o qual minimize "cost".

No processo seguinte, cada modo de predição é estimado utilizando um custo de distorção de taxa, em que maxCost é definido como o valor máximo o qual é inatingível como o valor estimado. Mais ainda, de modo a executar estimativas repetitivas, os melhores valores estimados nas condições individuais abaixo explicadas são respectivamente indicados por minCost, minCost1, e minCost2, e os índices de modo de predição atribuídos a estes são respectivamente indicados por best\_mode, best\_mode1, e best\_mode2.

As variáveis minCost e best\_mode representam o melhor valor

estimado e o índice de modo de predição correspondente quando predizendo todos os pixels no bloco relevante; as variáveis `minCost1` e `best_mode1` representam o melhor valor estimado e o índice de modo de predição correspondente quando predizendo somente os pixels cujo `Ref` correspondente tem um valor de 0; e as variáveis `minCost2` e `best_mode2` representam o melhor valor estimado e o índice de modo de predição correspondente quando predizendo somente os pixels cujo `Ref` correspondente tem um valor de 1.

Após inicializar cada um de `minCost`, `minCost1`, e `minCost2` em `maxCost` (ver etapa B1), uma imagem predita para o modo de predição `pmode` é gerada para cada pixel (ver etapas B2 até B8).

Na geração de imagem predita, após o índice de pixel "`pix`" no bloco relevante ser inicializado para zero (ver etapa B2), o seguinte processo (etapa B3 até B6) é repetidamente executado enquanto `pix` é incrementado por 1 (ver etapa B7), até `pix` atingir o número "`maxPix`" de pixels no bloco (ver etapa B8), por meio disto determinando um modo de predição o qual minimiza "`cost`".

Primeiro, de acordo com `Ref[pix]`, é determinado se existe ou não uma imagem compensada em disparidade no pixel correspondente (ver etapa B3).

Se nenhuma imagem compensada em disparidade estiver presente, a chave de referência 109 é operada de modo que o codificador preditivo de imagem diferencial 105 refira à memória de imagem de câmera decodificada 107 (ver etapa B4). Em contraste, se existir uma imagem compensada em disparidade, a chave de referência 109 é operada de modo que o codificador preditivo de imagem diferencial 105 refira à memória de imagem diferencial decodificada 108 (ver etapa B5).

Então, pela utilização da imagem de referência ajustada, um valor predito `Pred[pix]` para o pixel relevante é computado no modo de predição `pmode` (ver etapa B6).

Após a geração de imagem predita ser completada para todos os pixels no bloco, três tipos da soma de diferenças absolutas entre a ima-

gem predita e a imagem original são computados pelas seguintes fórmulas.

$$SAD1 = \sum |Sub[Pix] - Pred[Pix]| \cdot (1 - Ref[Pix])$$

$$SAD2 = \sum |Sub[Pix] - Pred[Pix]| \cdot Ref[Pix]$$

$$SAD = SAD1 + SAD2$$

5 em que SAD1 indica a soma das diferenças absolutas para os pixels os  
quais cada um não tem uma imagem compensada em disparidade; SAD2  
indica a soma das diferenças absolutas para os pixels os quais cada um tem  
uma imagem compensada em disparidade; e SAD indica a soma das dife-  
renças absolutas para todos os pixels no bloco. Além disso,  $\Sigma$  indica compu-  
10 tar a soma total para  $pix=0$  até  $maxPix-1$ . Apesar das fórmulas acima em-  
pregarem operações de multiplicação, a computação da soma de diferenças  
absolutas pode ser implementada utilizando uma simples ramificação na  
condição porque os pixels para computar SAD1 e SAD2 são exclusivos um  
para o outro.

15 Pela utilização de SAD, SAD1 E SAD2 acima, e pmode, os cus-  
tos de distorção de taxa denominados cost, cost1 e cost2 são computados  
de acordo com a fórmula acima descrita (ver etapa B10).

Os resultados computados são respectivamente comparados  
com os melhores custos de distorção de taxa os quais já foram computados  
20 (ver etapas B11, B13, e B15). Para cada caso, se o modo de predição pmo-  
de puder realizar um custo menor, a variável a qual indica o melhor modo de  
predição e o melhor valor de custo são atualizadas (ver etapas B12, B14, e  
B16).

Após a estimativa de todos os modos de predição ser completa-  
25 da, é determinado se diferentes modos de predição estão atribuídos a cada  
pixel que tem uma imagem compensada em disparidade e cada pixel que  
não tem compensação de disparidade, ou o mesmo modo de predição é a-  
tribuído a todos os pixels (ver etapas B19 a B21).

Primeiro, é determinado se best\_mode, best\_mode1, e  
30 best\_mode2 são todos os mesmos (ver etapa B19).

Se todos os elementos acima forem os mesmos, isto indica que  
um único modo pode ser utilizado no bloco, e assim best\_mode é armaze-

nado como o modo de predição PMODE utilizado na codificação preditiva (ver etapa B22). A operação é então terminada.

Se pelo menos um dos três indicar um modo diferente, então a soma de minCost1, minCost2, e OHCost é comparada a minCost (ver etapa  
5 B20).

Aqui, a primeira (soma) indica o custo requerido quando diferentes modos são utilizados no bloco relevante, e OHCost representa um custo de extra em consideração que dois modos de predição são codificados.

Se a primeira prover um melhor custo, o conjunto de  
10 best\_mode1, e best\_mode2 é armazenado como PMODE (ver etapa B21). A operação é então terminada.

Se a última prover um melhor custo, best\_mode é armazenado como PMODE (ver etapa B22), e a operação é terminada.

Na presente modalidade, se um modo de predição é utilizado ou  
15 dois modos de predição são utilizados no bloco é selecionado.

No entanto, um modo de predição pode sempre ser utilizado. Em tal caso, somente "cost" é computado na etapa B10; as etapas B13 até B16 e B19 até B22 são omitidas; e PMODE pode ser utilizado ao invés de best\_mode.

20 Por outro lado, dois modos de predição podem sempre ser utilizados. Em tal caso, a computação de SAD na etapa B9, a computação de Cost na etapa B10, e as etapas B11 até B12, B19 até B20, e B22 são desnecessárias; e se o resultado da determinação na etapa B18 for SIM, a operação prossegue para a etapa B21.

25 O modo de predição na presente modalidade indica como prever o valor de pixel.

Como o método de predição, existe a predição intraquadro para prever o valor de pixel utilizando um bloco periférico o qual pertence ao mesmo quadro que o presente bloco e já foi codificado e decodificado, e a  
30 predição interquadros para prever o valor de pixel utilizando outro quadro o qual já foi codificado e decodificado, e os modos de predição acima descritos incluem ambos os métodos. Um vetor de movimento requerido na predição

interquadros também funciona como um modo de predição. Além disso, qualquer método o qual possa predizer o valor de pixel utilizando dados os quais já foram codificados e decodificados pode funcionar como um modo de predição.

5                   A Figura 4 mostra um aparelho de decodificação de vídeo 200 como uma modalidade da presente invenção.

O aparelho de decodificação de vídeo 200 inclui uma unidade de entrada de dados codificados 201 para inserir os dados codificados no aparelho; uma unidade de entrada de imagem de referência 202 para inserir  
10                   uma imagem decodificada da câmera A no aparelho; uma unidade de entrada de informações de disparidade 203 para inserir as informações de disparidade, as quais indicam para qual posição sobre a imagem como o alvo de decodificação cada pixel na imagem decodificada da câmera A corresponde, no aparelho; um gerador de imagem compensada em disparidade 204 para  
15                   gerar uma imagem compensada em disparidade na posição da câmera B; uma unidade de ajuste de imagem de referência 205 para atribuir um grupo de imagens de referência para cada pixel com base no estado da imagem compensada em disparidade; uma memória de imagem de câmera decodificada 206 para armazenar as imagens decodificadas (como um grupo de i-  
20                   magens de referência) de imagens obtidas pela câmera B; uma memória de imagem diferencial decodificada 207 para armazenar as imagens decodificadas (como outro grupo de imagens de referência) das imagens diferenci-  
ais; uma chave de referência 208 para trocar a memória de imagem de referência com base nas informações de ajuste sobre o grupo de imagens de  
25                   referência; e um decodificador de imagem diferencial para decodificar os dados codificados inseridos.

As Figuras 5 e 6 são fluxogramas executados pelo aparelho de decodificação de vídeo 200 que tem a estrutura acima, e mostram um fluxo de operação para decodificar um quadro de dados codificados da câmera B.  
30                   O fluxo de operação será abaixo explicado em detalhes.

Aqui, é assumido que um quadro da câmera A ao mesmo tempo em que o quadro de decodificação-alvo já foi decodificado, e as informações

de disparidade relevantes também foram obtidas.

Primeiro, os dados codificados da câmera B são inseridos na unidade de entrada de dados codificados 201 (ver etapa C1). Aqui, uma imagem decodificada da câmera A, a qual tem o mesmo tempo de exibição que a imagem de entrada da câmera B, é inserida pela unidade de entrada de imagem de referência 202.

A seguir, uma imagem compensada em disparidade é gerada no gerador de imagem compensada em disparidade 204 com base na imagem decodificada da câmera A e nas informações de disparidade (ver etapa C2). O processo nesta etapa é igual àquele executado na etapa A2 acima explicada na Figura 2.

Os dados codificados inseridos de cada bloco são decodificados utilizando a imagem compensada em disparidade, de modo que uma imagem decodificada da câmera B é obtida (ver etapas C3 até C9).

Neste processo, "blk" é um índice para cada bloco, e maxBlk indica o número total de blocos em uma imagem. Após blk ser inicializado para zero (ver etapa C3), o processo seguinte (etapas C4 até C7) é repetidamente executado enquanto blk é incrementado por 1 (ver etapa C12), até blk atingir o número total de blocos maxBlk (ver etapa C9).

Primeiro, a unidade de ajuste de imagem de referência 205 gera as informações Ref para cada pixel no bloco blk de acordo com um método similar àquele na etapa A6 acima descrita, em que Ref indica uma referência utilizada quando a imagem diferencial foi sujeita a uma codificação preditiva (ver etapa C4).

Os dados codificados de entrada são decodificados no decodificador de imagem diferencial 209 pela utilização das informações acima, de modo que cada valor decodificado Dec da imagem diferencial seja obtido (ver etapa C5).

O valor decodificado Dec é diretamente utilizado para decodificar um quadro em outro tempo ou outro bloco, e é assim armazenado na memória de imagem diferencial decodificada 207. (ver etapa C6).

Além disso, uma imagem decodificada da imagem obtida pela

câmera B é adquirida pela computação da soma de Dec e da imagem compensada em disparidade Syn para cada pixel no bloco blk.

A imagem decodificada adquirida é emitida do aparelho de decodificação de vídeo 200, e é simultaneamente armazenada na memória de  
 5 imagem de câmera decodificada 206. (ver etapa C7).

A Figura 6 mostra um fluxo de operação detalhado do processo (na etapa C5) de decodificação da imagem diferencial para cada bloco, o qual é executado no decodificador de imagem diferencial 209.

Primeiro, as informações de modo de predição são decodificadas dos dados codificados, e são ajustadas como pmode[0] (ver etapa D1).  
 10 Neste processo, se outro item de informações de modo de predição for incluído nos dados codificados (ver etapa D2), o item é também decodificado e ajustado como pmode[1] (ver etapa D3). Se nenhum outro item de informações de modo de predição estiver incluído (etapa D2), pmode[1] é ajustado  
 15 para o mesmo valor de pmode[0] (ver etapa D4).

A seguir, o resíduo de predição para cada pixel, o qual está incluído nos dados codificados, é decodificado e ajustado como Dres (ver etapa D5).

A operação após D5 é executada para cada pixel no bloco. Isto  
 20 é, após o índice de pixel "pix" ser inicializado para zero (ver etapa D6), o seguinte processo (etapas D7 até D11) é repetidamente executado enquanto pix é incrementado por 1 (ver etapa D12), até pix atingir o número "maxPix" de pixels no bloco (ver etapa D13).

Na operação aplicada a cada pixel, primeiro, a chave de referência  
 25 208 é operada de acordo com o valor de Ref[pix] (ver etapa D7), de modo que o armazenamento temporário de referência seja determinado para a memória de imagem de câmera decodificada 206 (ver etapa D8) ou a memória de imagem diferencial decodificada 207 (ver etapa D9).

A seguir, o valor Pred de pixel pix é predito no modo de predição  
 30 pmode[Ref[pix]] (ver etapa D10).

Então, o valor decodificado Dec[pix] da imagem diferencial é obtido pela soma de Dres[pix] e Pred (ver etapa D11).

Na presente modalidade, as informações de disparidade, as quais indicam para qual posição sobre a imagem da câmera B cada pixel sobre a imagem de câmera de referência corresponde, são obtidas fora do aparelho de codificação de vídeo 100 ou do aparelho de decodificação de vídeo 200, e as informações são codificadas, transmitidas, e decodificadas.

No entanto, as informações de disparidade podem ser obtidas e codificadas no aparelho de codificação de vídeo 100, e as informações codificadas podem ser emitidas juntamente com os dados codificados da imagem diferencial. Similarmente, as informações de disparidade podem ser recebidas no aparelho de decodificação de vídeo 200, e as informações recebidas podem ser decodificadas e utilizadas.

Além disso, as informações, as quais não indicam diretamente para qual posição sobre a imagem da câmera B cada pixel sobre a imagem de câmera de referência corresponde, mas provê a relação correspondente acima descrita devido a uma conversão aplicada nas informações, podem ser inseridas e convertidas para informações de disparidade (as quais indicam a relação correspondente) no aparelho de codificação de vídeo 100 e no aparelho de decodificação de vídeo 200.

Como um exemplo de tais informações, existe as informações que consistem em informações tridimensionais de um objeto de imagem formada e parâmetros de câmera. As informações tridimensionais podem ser coordenadas tridimensionais de cada parte do objeto de imagem formada, ou podem indicar uma distância do objeto de imagem formada para a câmera relevante.

Além disso, nas modalidades, a troca do grupo de imagens de referência como o alvo de referência pode ser executada para cada pixel na imagem de câmera-alvo de codificação. No entanto, a troca pode ser executada para uma unidade de seção específica a qual inclui uma pluralidade de pixels. Por exemplo, quando a troca pode ser executada para cada bloco que tem um tamanho específico, um alvo de referência adequado para um número de pixels no bloco pode ser selecionado, ou o grupo de imagens de câmera decodificado pode ser selecionado se o bloco incluir pelo menos um

pixel cujo valor de pixel da imagem compensada em disparidade for ineficaz. Uma seleção baseada em uma unidade de seção apropriada pode reduzir o tempo de processamento.

- Os processos de codificação e de decodificação de vídeo como
- 5 acima descritos podem ser implementados utilizando um recurso de hardware ou de software, e podem também ser implementados utilizando um computador e um programa de software. O programa de computador pode ser provido armazenando-o em um meio de armazenamento legível por computador, ou por meio de uma rede.

#### 10 APLICABILIDADE INDUSTRIAL

- De acordo com a presente invenção, quando sujeitando uma imagem diferencial obtida por compensação de disparidade para codificação preditiva, nenhuma nova informação adicional é utilizada, e uma imagem de referência apropriada é utilizada através de uma operação de troca para cada pixel. Portanto, o resíduo para o alvo de codificação pode ser reduzido
- 15 sem aumentar a quantidade de código requerida para as informações adicionais, por meio disto implementando uma codificação de vídeo eficiente da imagem de vídeo de múltiplos pontos de vista inteira.

## REIVINDICAÇÕES

1. Método de codificação de vídeo para codificar uma imagem de vídeo de múltiplos pontos de vista pela utilização de compensação de disparidade, o método compreendendo as etapas de:

- 5                    gerar, com base nas informações de disparidade entre uma imagem de câmera de referência, a qual é obtida pela decodificação de uma imagem de câmera já codificada, e uma imagem de câmera-alvo de codificação que corresponde à imagem de câmera de referência, uma imagem compensada em disparidade para a imagem de câmera-alvo de codificação;
- 10                  ajustar as informações de alvo de referência, as quais indicam um grupo de imagens de referência como o alvo de referência, para cada unidade de seção predeterminada sobre a imagem de câmera-alvo de codificação de acordo com um estado da imagem compensada em disparidade;  
                      selecionar, para cada unidade de seção predeterminada sobre a
- 15                  imagem de câmera-alvo de codificação, um de um grupo de imagens diferenciais decodificadas, o qual é obtido pela decodificação de uma imagem diferencial entre a imagem de câmera já codificada e a imagem compensada em disparidade, e um grupo de imagens de câmera decodificadas, o qual é obtido pela decodificação da imagem de câmera já codificada como o alvo
- 20                  de referência com base nas informações de alvo de referência ajustadas;  
                      codificar preditivamente uma imagem diferencial entre a imagem de câmera-alvo de codificação e a imagem compensada em disparidade referindo a uma imagem de referência incluída no grupo de imagens selecionado como o alvo de referência;
- 25                  armazenar uma imagem diferencial decodificada, a qual é obtida pela decodificação da imagem diferencial codificada, como um constituinte do grupo de imagens diferenciais decodificadas; e  
                      armazenar uma imagem de câmera decodificada, a qual é obtida pela decodificação da imagem de câmera já codificada com base na imagem
- 30                  diferencial decodificada, como um constituinte do grupo de imagens de câmera decodificadas.

2. Método de codificação de vídeo de acordo com a reivindica-

ção 1, em que:

a etapa de codificar preditivamente a imagem diferencial inclui computar um custo de codificação que ocorre quando cada um dos métodos de predição predeterminados é aplicado a cada bloco de processamento de codificação o qual pertence à imagem diferencial e consiste em uma pluralidade de pixels, em que a predição de vídeo de um bloco de processamento de codificação é executada utilizando um único método de predição o qual provê o custo de codificação mínimo.

3. Método de codificação de vídeo de acordo com a reivindicação 1, em que:

a etapa de codificar preditivamente a imagem diferencial inclui computar um custo de codificação que ocorre quando cada um dos métodos de predição predeterminados é aplicado a cada grupo de pixels em cada bloco de processamento de codificação o qual pertence à imagem diferencial e consiste em uma pluralidade de pixels, em que cada grupo de pixels tem uma imagem de referência individual, e a codificação é executada enquanto selecionando se a predição de vídeo de um bloco de processamento de codificação é executada utilizando um único método de predição ou uma pluralidade de métodos de predição, de acordo com os custos computados.

4. Método de codificação de vídeo de acordo com a reivindicação 1, em que a unidade de seção predeterminada é um pixel.

5. Método de decodificação de vídeo para decodificar uma imagem de vídeo de múltiplos pontos de vista pela utilização de compensação de disparidade, o método compreendendo as etapas de:

gerar, com base nas informações de disparidade entre uma imagem de câmera de referência já decodificada e uma imagem de câmera-alvo de decodificação que corresponde à imagem de câmera de referência, uma imagem compensada em disparidade para a imagem de câmera-alvo de decodificação;

ajustar as informações de alvo de referência, as quais indicam um grupo de imagens de referência como o alvo de referência, para cada unidade de seção predeterminada sobre a imagem de câmera-alvo de deco-

dificação de acordo com um estado da imagem compensada em disparidade;

5                   selecionar, para cada unidade de seção predeterminada sobre a imagem de câmera-alvo de decodificação, um de um grupo de imagens diferenciais decodificadas que tem a imagem diferencial entre a imagem de câmera já decodificada e a imagem compensada em disparidade, e um grupo de imagens de câmera decodificadas que tem a imagem de câmera já decodificada, como o alvo de referência com base nas informações de alvo de referência ajustadas;

10                  decodificar uma imagem diferencial entre a imagem de câmera-alvo de decodificação e a imagem compensada em disparidade referindo a uma imagem de referência incluída no grupo de imagens selecionado como o alvo de referência, com base em um resíduo de predição de cada pixel, o qual é obtido pela decodificação de dados codificados de entrada;

15                  armazenar a imagem diferencial decodificada como um constituinte do grupo de imagens diferenciais decodificadas; e

                  adicionar a imagem diferencial decodificada à imagem compensada em disparidade, emitindo a soma destas como uma imagem de câmera decodificada da imagem de vídeo de múltiplos pontos de vista, e armazenar  
20                  a imagem de câmera decodificada como um constituinte do grupo de imagens de câmera decodificadas.

                  6. Método de decodificação de vídeo de acordo com a reivindicação 5, em que:

                  a etapa de decodificar a imagem diferencial inclui decodificar a  
25                  imagem diferencial em um modo tal que um único método de predição designado é aplicado a cada bloco de processamento de decodificação o qual pertence à imagem de câmera-alvo de decodificação e consiste em uma pluralidade de pixels.

                  7. Método de decodificação de vídeo de acordo com a reivindicação 5, em que:

                  a etapa de decodificar a imagem diferencial inclui decodificar a imagem diferencial enquanto selecionando para cada bloco de processa-

mento de decodificação o qual pertence à imagem de câmera-alvo de decodificação e consiste em uma pluralidade de pixels, se um único método de predição designado ou uma pluralidade de métodos de predição designados são utilizados.

5                   8. Método de decodificação de vídeo de acordo com a reivindicação 5, em que a unidade de seção predeterminada é um pixel.

9. Aparelho de codificação de vídeo para codificar uma imagem de vídeo de múltiplos pontos de vista pela utilização de compensação de disparidade, o aparelho compreendendo:

10                   um dispositivo para gerar, com base nas informações de disparidade entre uma imagem de câmera de referência, a qual é obtida pela decodificação de uma imagem de câmera já codificada, e uma imagem de câmera-alvo de codificação que corresponde à imagem de câmera de referência, uma imagem compensada em disparidade para a imagem de câmera-alvo  
15 de codificação;

                    um dispositivo para ajustar as informações de alvo de referência, as quais indicam um grupo de imagens de referência como o alvo de referência, para cada unidade de seção predeterminada sobre a imagem de câmera-alvo de codificação de acordo com um estado da imagem compensada  
20 em disparidade;

                    um dispositivo para selecionar, para cada unidade de seção predeterminada sobre a imagem de câmera-alvo de codificação, um de um grupo de imagens diferenciais decodificadas, o qual é obtido pela decodificação de uma imagem diferencial entre a imagem de câmera já codificada e a imagem compensada em disparidade, e um grupo de imagens de câmera decodificadas, o qual é obtido pela decodificação da imagem de câmera já codificada como o alvo de referência com base nas informações de alvo de referência ajustadas;  
25

                    um dispositivo para codificar preditivamente uma imagem diferencial entre a imagem de câmera-alvo de codificação e a imagem compensada em disparidade referindo a uma imagem de referência incluída no grupo de imagens selecionado como o alvo de referência;  
30

um dispositivo para armazenar uma imagem diferencial decodificada, a qual é obtida pela decodificação da imagem diferencial codificada, como um constituinte do grupo de imagens diferenciais decodificadas; e

um dispositivo para armazenar uma imagem de câmera decodificada, a qual é obtida pela decodificação da imagem de câmera já codificada com base na imagem diferencial decodificada, como um constituinte do grupo de imagens de câmera decodificadas.

10. Aparelho de codificação de vídeo de acordo com a reivindicação 9, em que a unidade de seção predeterminada é um pixel.

10 11. Aparelho de decodificação de vídeo para decodificar uma imagem de vídeo de múltiplos pontos de vista pela utilização de compensação de disparidade, o método compreendendo as etapas de:

um dispositivo para gerar, com base nas informações de disparidade entre uma imagem de câmera de referência já decodificada e uma imagem de câmera-alvo de decodificação que corresponde à imagem de câmera de referência, uma imagem compensada em disparidade para a imagem de câmera-alvo de decodificação;

um dispositivo para ajustar as informações de alvo de referência, as quais indicam um grupo de imagens de referência como o alvo de referência, para cada unidade de seção predeterminada sobre a imagem de câmera-alvo de decodificação de acordo com um estado da imagem compensada em disparidade;

um dispositivo para selecionar, para cada unidade de seção predeterminada sobre a imagem de câmera-alvo de decodificação, um de um grupo de imagens diferenciais decodificadas que tem a imagem diferencial entre a imagem de câmera já decodificada e a imagem compensada em disparidade, e um grupo de imagens de câmera decodificadas que tem a imagem de câmera já decodificada, como o alvo de referência com base nas informações de alvo de referência ajustadas;

um dispositivo para decodificar uma imagem diferencial entre a imagem de câmera-alvo de decodificação e a imagem compensada em disparidade referindo a uma imagem de referência incluída no grupo de ima-

gens selecionado como o alvo de referência, com base em um resíduo de predição de cada pixel, o qual é obtido pela decodificação de dados codificados de entrada;

um dispositivo para armazenar a imagem diferencial decodificada como um constituinte do grupo de imagens diferenciais decodificadas; e

um dispositivo para adicionar a imagem diferencial decodificada à imagem compensada em disparidade, emitindo a soma destas como uma imagem de câmera decodificada da imagem de vídeo de múltiplos pontos de vista, e armazenar a imagem de câmera decodificada como um constituinte do grupo de imagens de câmera decodificadas.

12. Aparelho de decodificação de vídeo de acordo com a reivindicação 11, em que a unidade de seção predeterminada é um pixel.

13. Programa de codificação de vídeo pelo qual um computador executa um processo para implementar o método de codificação de vídeo como definido na reivindicação 1.

14. Programa de decodificação de vídeo pelo qual um computador executa um processo para implementar o método de decodificação de vídeo como definido na reivindicação 5.

15. Meio de armazenamento legível por computador o qual armazena um programa de codificação de vídeo pelo qual um computador executa um processo para implementar o método de codificação de vídeo como definido na reivindicação 1.

16. Meio de armazenamento legível por computador o qual armazena um programa de decodificação de vídeo pelo qual um computador executa um processo para implementar o método de decodificação de vídeo como definido na reivindicação 1.

FIG. 1

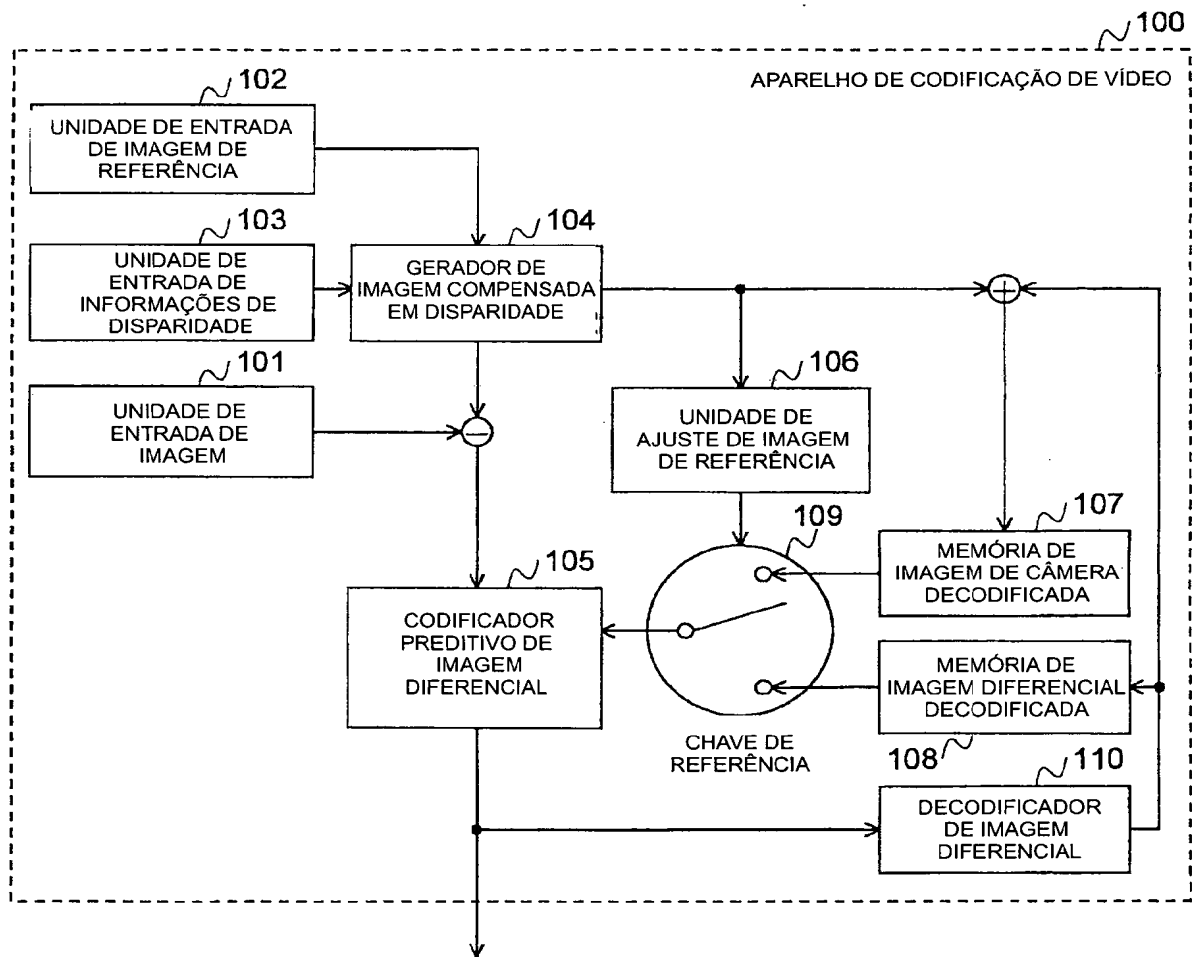


FIG. 2

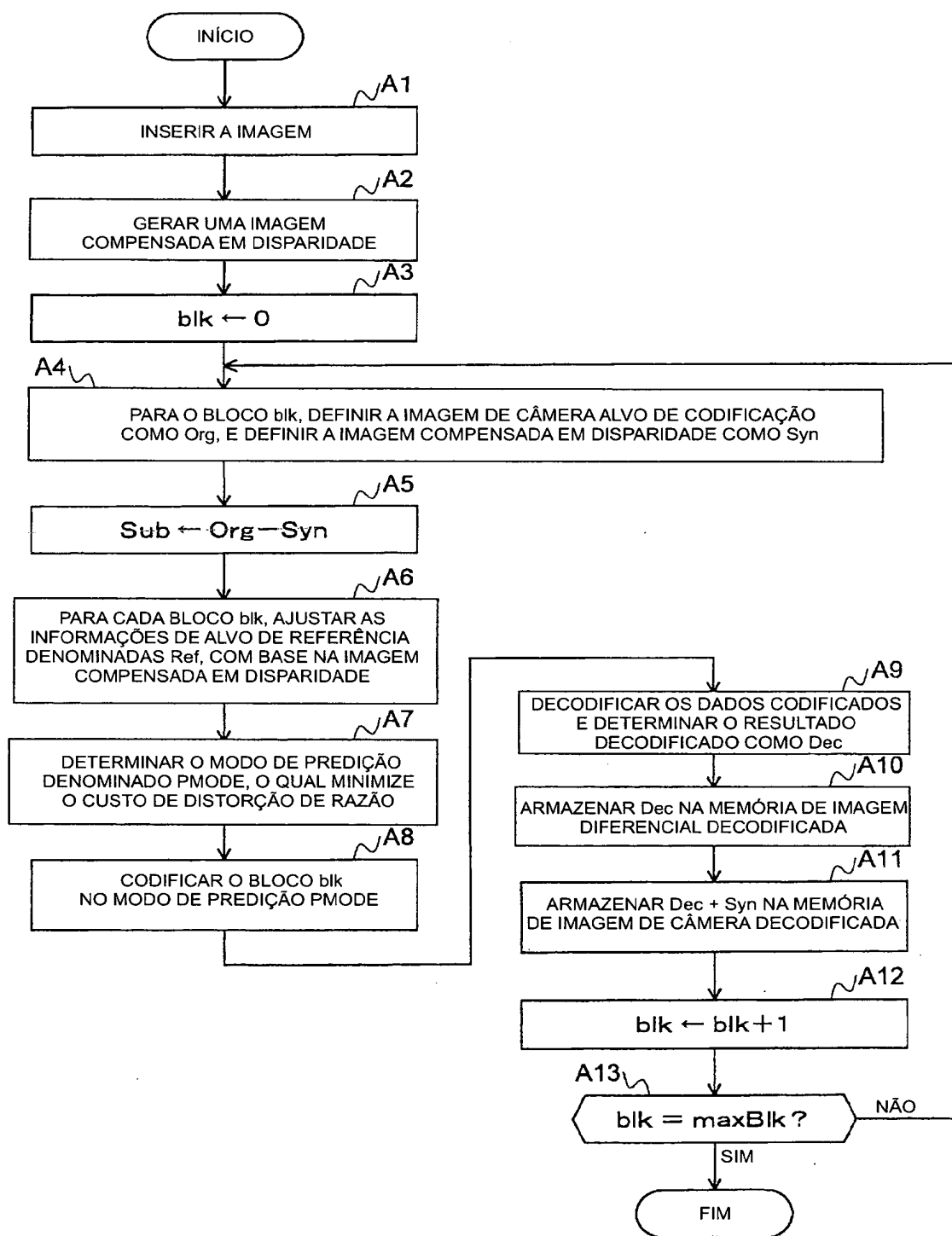


FIG. 3

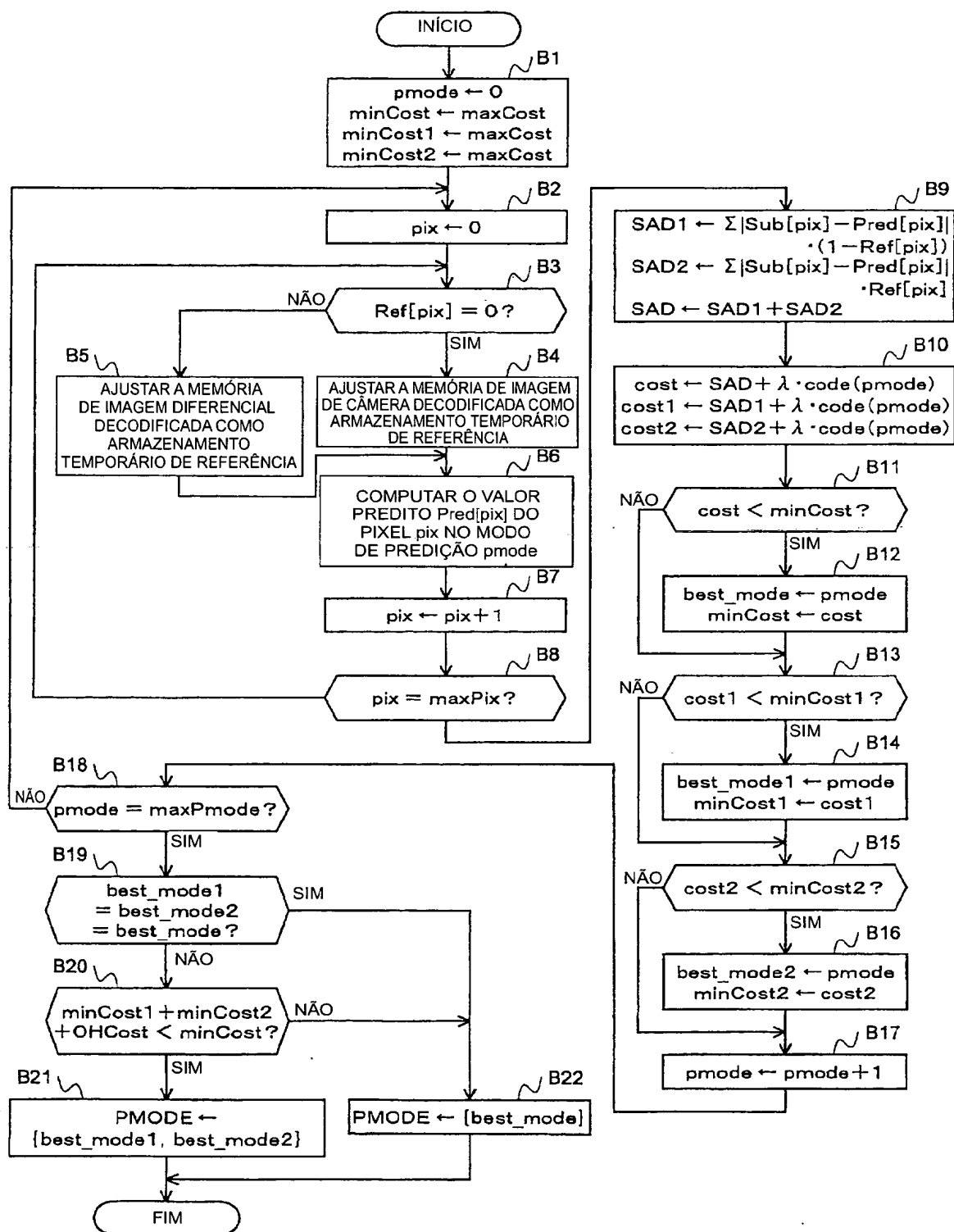


FIG. 4

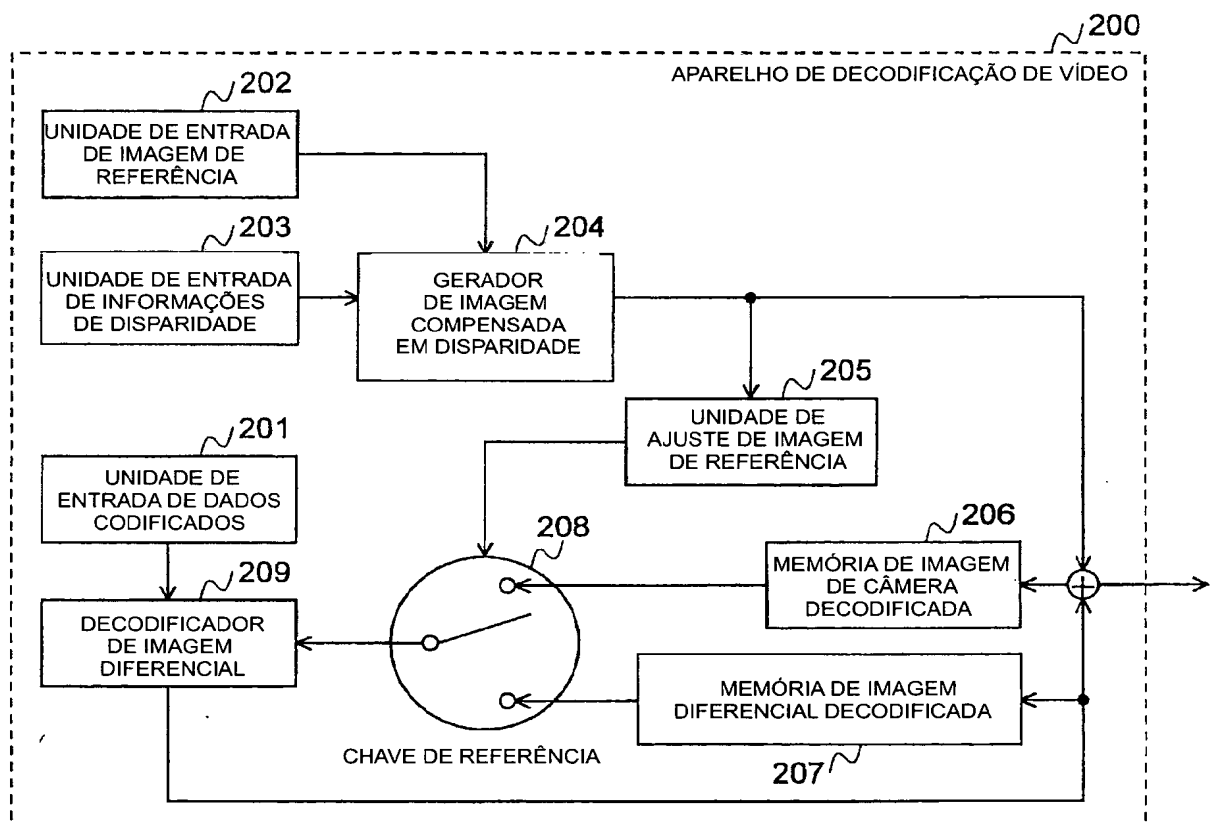


FIG. 5

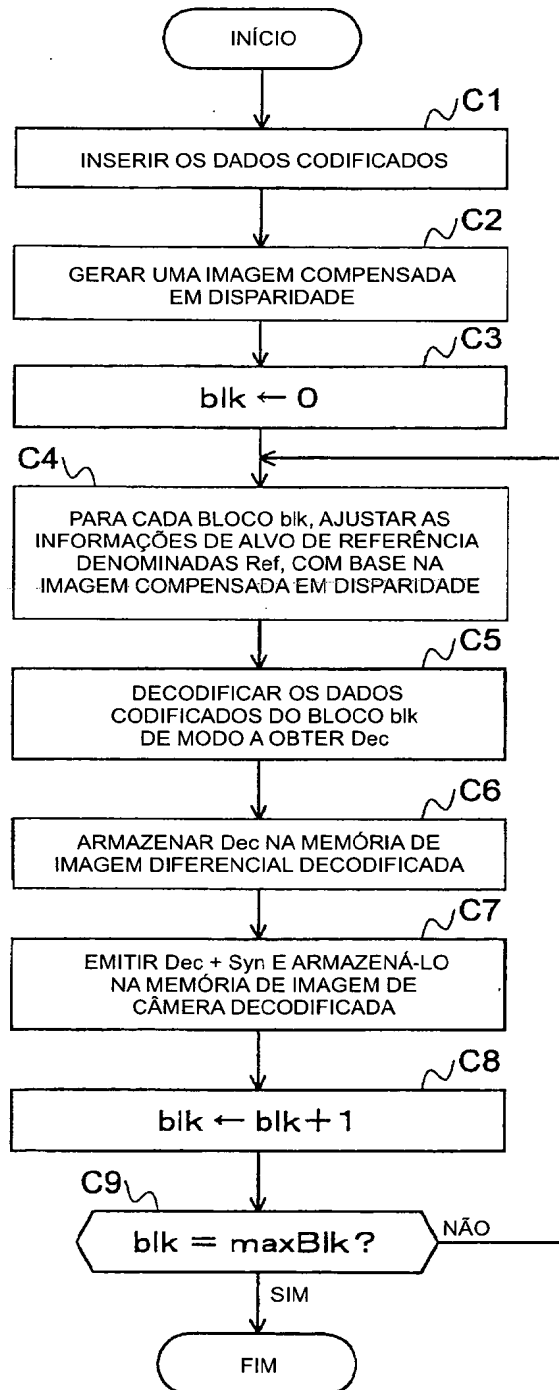


FIG. 6

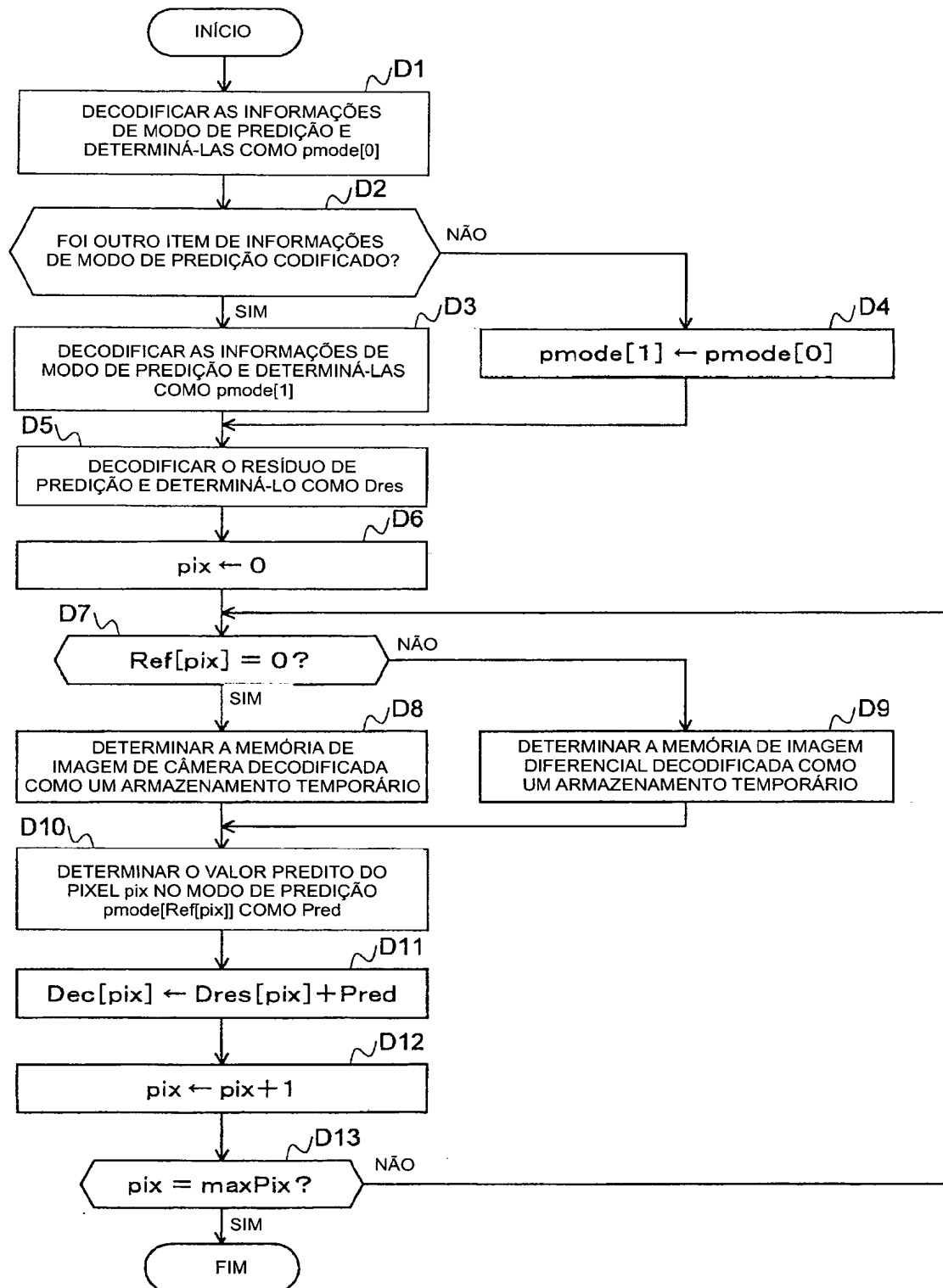


FIG. 7

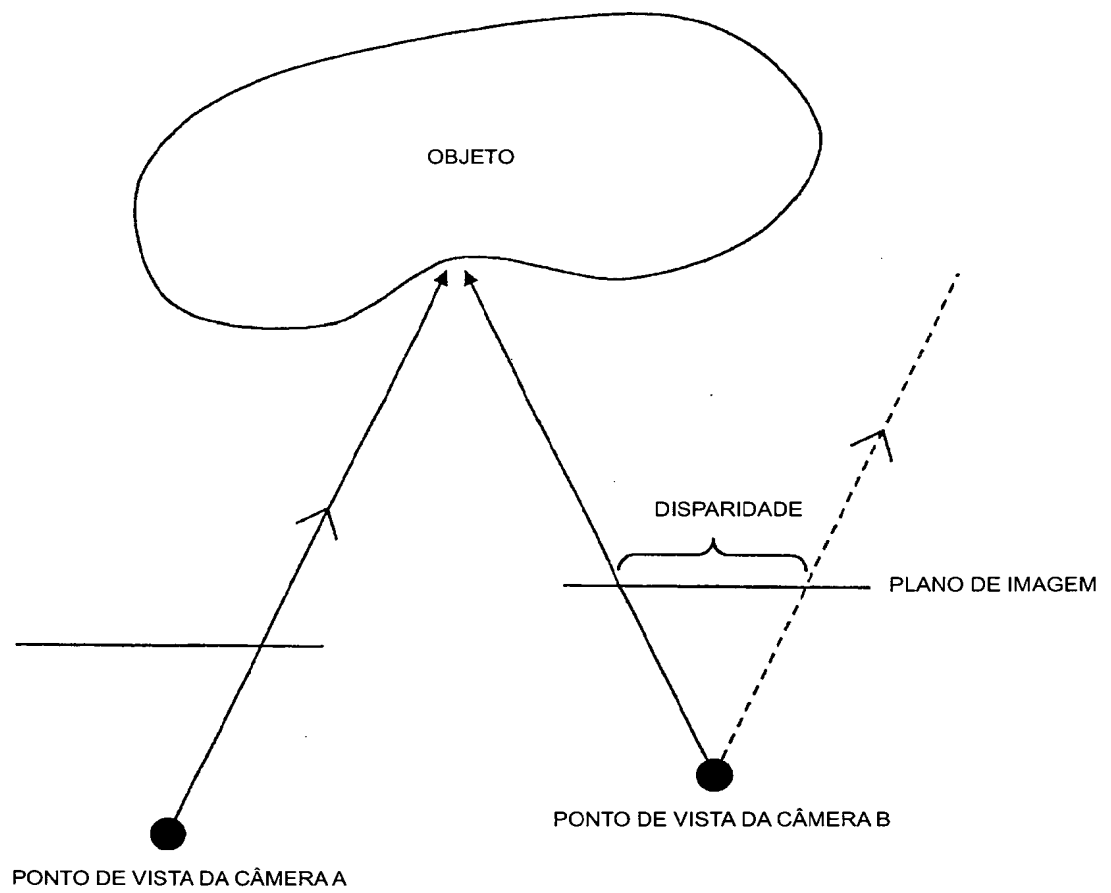


FIG. 8

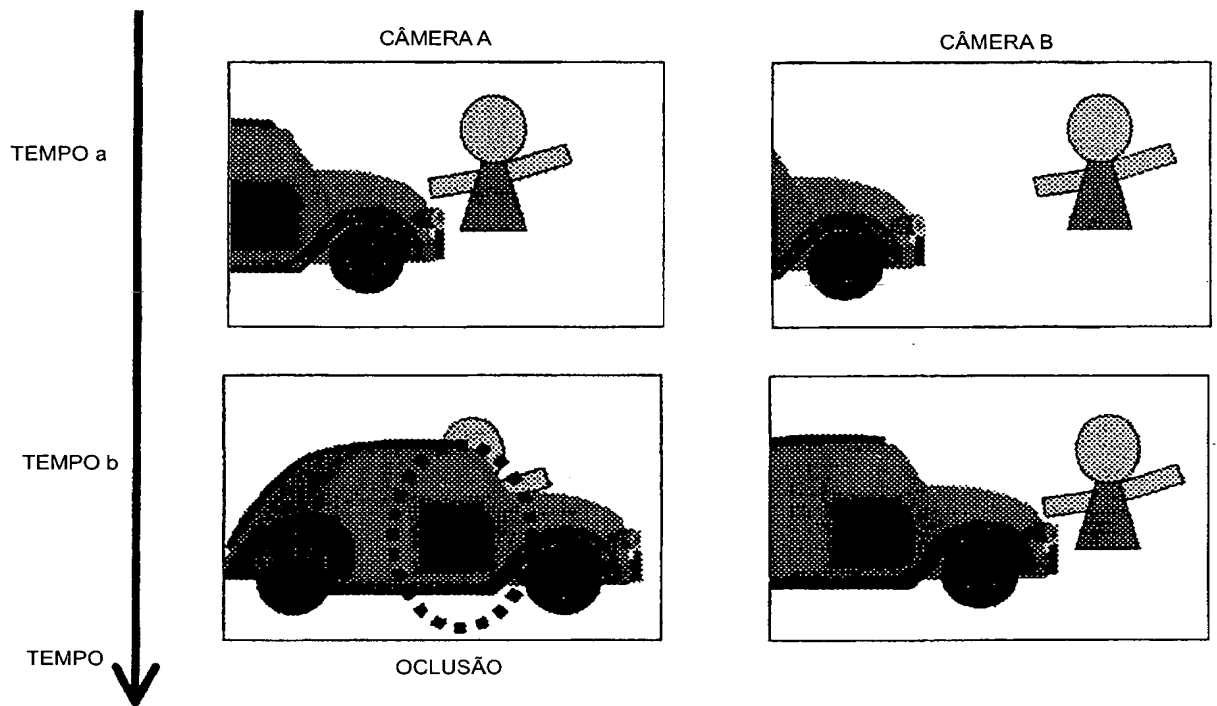
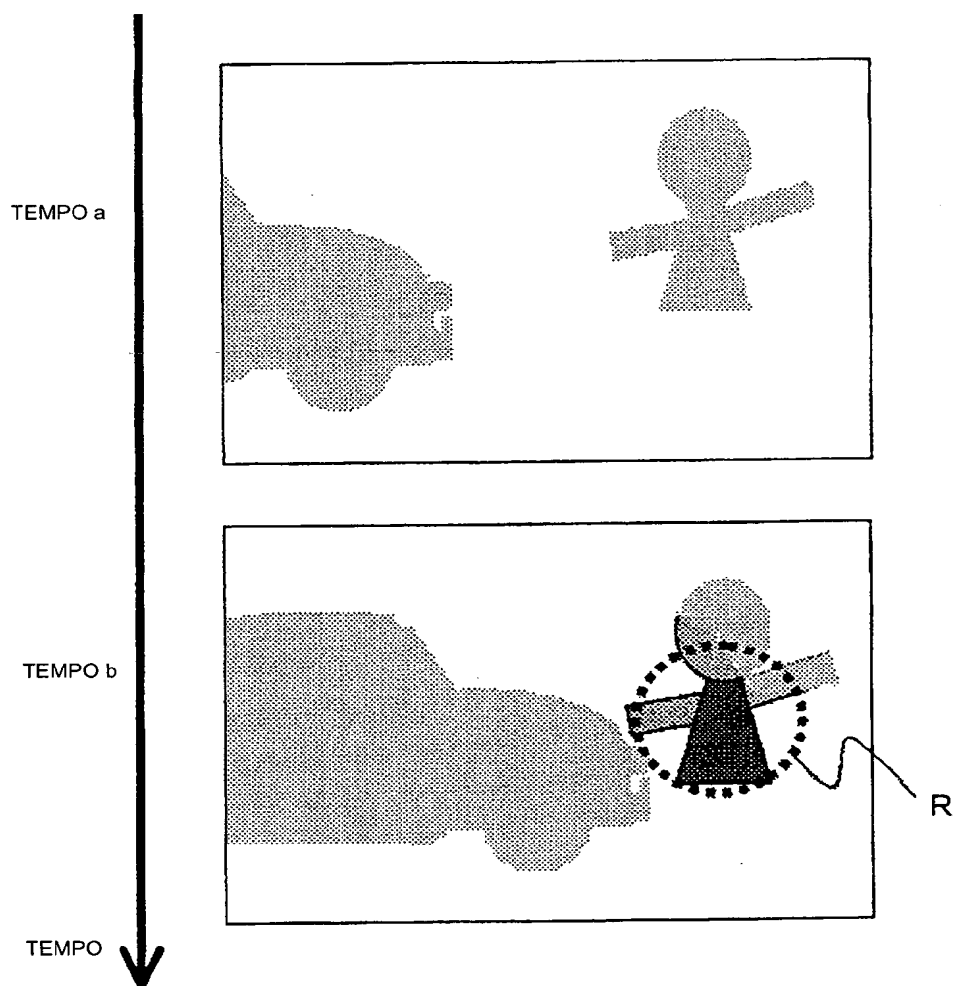


FIG. 9



## RESUMO

Patente de Invenção: **"MÉTODO DE CODIFICAÇÃO E MÉTODO DE DECODIFICAÇÃO DE VÍDEO, APARELHOS PARA OS MESMOS, PROGRAMAS PARA OS MESMOS, E MEIO DE ARMAZENAMENTO O QUAL**  
5 **ARMAZENA OS PROGRAMAS"**.

A presente invenção refere-se a um método de codificação de vídeo para gerar, com base nas informações de disparidade entre uma imagem de câmera de referência já codificada e uma imagem de câmera-alvo de codificação que corresponde à imagem de câmera de referência, uma  
10 imagem compensada em disparidade por meio de predição de imagem intercâmeras; e codificar uma imagem diferencial entre a imagem de câmera-alvo de codificação e a imagem compensada em disparidade. O método inclui selecionar, para cada unidade de seção predeterminada sobre a imagem diferencial, um de um grupo de imagens diferenciais decodificadas obtido  
15 pela decodificação de uma imagem diferencial entre a imagem de câmera já codificada e a imagem compensada em disparidade e um grupo de imagens de câmera decodificadas obtido pela decodificação da imagem de câmera já codificada, determinando se existe ou não uma imagem compensada em disparidade na posição correspondente, isto é, se o pixel correspondente na  
20 imagem compensada em disparidade tem ou não um valor efetivo.