



República Federativa do Brasil
Ministério da Economia
Instituto Nacional da Propriedade Industrial

(11) PI 1101662-0 B1

(22) Data do Depósito: 07/04/2011

(45) Data de Concessão: 26/10/2021



(54) Título: MÉTODO PARA PROCESSAR VÍDEO DE UMA SEQUÊNCIA DE PARES DE IMAGENS ESTEREOSCÓPICAS

(51) Int.Cl.: H04N 19/577; H04N 13/00; H04N 19/51; H04N 19/597.

(52) CPC: H04N 19/577; H04N 13/0003; H04N 19/51; H04N 19/597.

(30) Prioridade Unionista: 09/04/2010 FR 1052693.

(73) Titular(es): INTERDIGITAL MADISON PATENT HOLDINGS.

(72) Inventor(es): JONATHAN KERVEC; HASSANE GUERMOUD; EMMANUEL JOLLY.

(57) Resumo: MÉTODO PARA PROCESSAR IMAGENS ESTEREOSCÓPICAS E DISPOSITIVO CORRESPONDENTE. A invenção se refere a um método para processamento de vídeo de pelo menos uma imagem de uma sequência de vídeos, a dita sequência de vídeos compreendendo uma pluralidade de pares de imagens, cada par de imagem compreendendo uma primeira imagem (14,15, 16) e uma segunda imagem (17, 18, 19), as ditas primeira (14, 15, 16) e segunda (17, 18, 19) imagens sendo intencionadas para formar uma imagem estereoscópica. Para reduzir os defeitos de exibição, o método compreende a etapa de gerar pelo menos uma terceira imagem (17i, 18i, 19i) por interpolação temporal com compensação de movimento de pelo menos duas das ditas imagens secundárias (17,18,19).

"MÉTODO PARA PROCESSAR VÍDEO DE UMA SEQUÊNCIA DE PARES DE IMAGENS ESTEREOSCÓPICAS"

1. CAMPO DA INVENÇÃO

A invenção se refere ao campo de processamento de imagem ou vídeo, e, mais especificamente, ao processamento de imagens e/ou vídeo tridimensional. A invenção também se refere ao campo de interpolação de imagens com compensação de movimento.

2. TÉCNICA ANTERIOR

De acordo com a técnica anterior, um dos métodos usados em processamento de vídeo, para restaurar uma percepção de relevo, é a estereoscopia. De acordo com esse método, duas vistas de uma mesma cena são registradas, com duas diferentes câmeras de vídeo ou duas diferentes câmeras de quadro único, de dois diferentes pontos de vista lateralmente deslocados entre si. A captura de uma mesma cena por duas diferentes câmeras é mostrada na Figura 1, de acordo com uma concretização particular. A cena filmada por duas câmeras, chamadas a câmera esquerda G 11 e a câmera direita D 12, é composta de dois cilindros, o primeiro 100 desses cilindros sendo estático e o segundo 110 desses cilindros deslocando-se, de acordo com um movimento horizontal indo da esquerda para a direita, para captar três posições espaciais sucessivas nos tempos T1, T2 e T3. As câmeras esquerda G11 e direita D 12 são espaçadas entre si de acordo com uma distância predefinida (por exemplo, 6,5 cm, a distância correspondente à distância média separando os olhos de uma pessoa) e registram a mesma cena, de acordo com dois diferentes pontos de vista, respectivamente, 111 e 121. Cada uma das câmeras G 11 e D 12 filmam uma sequência de vídeos, composta de três imagens, cada uma das três imagens sendo capturada, respectivamente, nos tempos T1, T2 e T3. A aquisição da sequência de vídeos é representada no diagrama de sincronização 13 da Figura 1. A câmera G 11 registra as três imagens sucessivas 14, 15 e 16, respectivamente, e os tempos T1, T2 e T3, e a câmera D 12 registra as três imagens sucessivas 17, 18 e 19, respectivamente, nos mesmos tempos T1, T2 e T3, como, respectivamente, das imagens 14, 15 e 16. As imagens 14, 15 e 16 vão ser chamadas imagens esquerdas no restante da descrição, e as imagens 17, 18 e 19 vão ser chamadas imagens direitas. As imagens esquerda 14 e direita 17 são capturadas de um modo sincronizado e são representativas dos cilindros 100 e 110, no tempo T1, as imagens esquerda 15 e direita 18 são capturadas de um modo sincronizado e são representativas dos cilindros 100 e 110, no tempo T2, e as imagens esquerda 16 e direita 19 são capturadas de um modo sincronizado e são representativas dos cilindros 100 e 110, no tempo T3. As imagens esquerda 14 e direita 17 sendo capturadas por câmeras deslocadas espacialmente 11 e 12, a posição dos cilindros 100 e 110, representadas na imagem esquerda 14 pelos cilindros 100g e 110g, não é idêntica à posição dos mesmos cilindros 100 e 110, representados na imagem direita 17 pelos cilindros 100d e 110d, as duas imagens 14 e 17 sendo de dimensões idênti-

cas (por exemplo, 1.920 x 1.080 pixels em formato HD - de alta definição). As diferenças entre as posições dos cilindros 100g e 110g, em um lado, e as posições dos cilindros 100d e 110d, no outro lado, também aparecem no par de imagens 15 e 18 tomadas no tempo T2 e no par de imagens 16 e 19 tomadas no tempo T3.

5 Cada par de imagens esquerda e direita, isto é, as imagens esquerda 14 e direita 17 no tempo T1, as imagens esquerda 15 e direita 18 no tempo T2 e imagens esquerda 16 e direita 19 no tempo T3, é intencionado para formar uma imagem estereoscópica restaurando uma sintetização tridimensional da cena 10, nos tempos T1, T2 e T3, pelas imagens 14 a 19, é capturado em duas dimensões. Há várias técnicas para exibir os pares de imagens 10 14/17, 15/18 e 16/19, que propiciam a restauração de imagens estereoscópicas. Uma dessas técnicas consiste em apresentar, alternativamente, uma imagem esquerda, depois uma imagem direita, por exemplo, de acordo com a seguinte sequência temporal: 14 - 17 - 15 - 18 - 16 - 19. Essa apresentação alternada de imagens esquerda e direita no tempo é conhecida como apresentação sequencial temporal, um usuário desejando ver a sintetização em 15 DIAFRAGMA DE PRESSÃO de imagens da sequência tendo que usar óculos específicos, por exemplo, óculos ativos, para os quais a oclusão do olho esquerdo e do olho direito é, respectivamente, sincronizada com a exibição de imagens esquerda e direita em, por exemplo, dispositivo de exibição do tipo LCD ou plasma.

A exibição sequencial temporal, isto é, a exibição sucessiva de uma imagem esquerda, depois uma imagem direita, acarreta defeitos de exibição, devido ao fato de que as 20 imagens esquerda e direita, exibidas sucessivamente no tempo, foram capturadas de um modo sincronizado, quando áreas de imagens capturadas revelam movimentos.

3. RESUMO DA INVENÇÃO

A finalidade da invenção é superar pelo menos uma dessas desvantagens da técnica anterior. 25

Mais especificamente, a finalidade da invenção é reduzir acentuadamente os defeitos de exibição estereoscópica de sequências compreendendo os objetos em movimento.

A invenção se refere a um método para processamento de vídeo de pelo menos uma sequência de vídeos, a sequência de vídeos compreendendo uma pluralidade de pares 30 de imagens, cada par de imagens compreendendo uma primeira imagem e uma segunda imagem, as primeira e segunda imagens sendo intencionadas para formar uma imagem estereoscópica. O método compreende uma etapa de geração de pelo menos uma terceira imagem por interpolação temporal de movimento compensado de pelo menos duas das ditas imagens secundárias.

35 De acordo com uma característica particular, o método compreende uma etapa de estimativa de um vetor de movimento para cada pixel de uma pluralidade de pixels da terceira imagem, de pelo menos duas imagens secundárias.

Vantajosamente, o método compreende uma etapa de exibição de imagens primárias e terciárias em um dispositivo de exibição, as imagens primárias e terciárias sendo exibidas, respectivamente, sequencialmente.

De acordo com uma característica específica, o método compreende uma etapa de
5 geração de pelo menos uma quarta imagem por interpolação temporal de movimento compensado de pelo menos duas imagens primárias.

Vantajosamente, o método compreende uma etapa de exibição de imagens terciárias e quaternárias em um dispositivo de exibição, as imagens terciárias e quaternárias sendo, respectivamente, exibidas sequencialmente.

10 4. LISTA DE FIGURAS

A invenção vai ser melhor entendida e outros aspectos e vantagens específicos vão surgir mediante leitura da descrição apresentada a seguir, a descrição fazendo referência aos desenhos anexados, em que:

- a Figura 1, descrita acima, mostra a captura de imagens de uma sequência de vídeos, representativa de uma cena 10 por dois dispositivos de registro, de acordo com uma concretização particular da técnica anterior;

- a Figura 2 mostra um método para registro temporal de imagens capturadas de acordo com o método descrito na Figura 1, de acordo com uma concretização da invenção;

- a Figura 3 mostra um método para interpolação temporal de imagens capturadas
20 de acordo com o método descrito na Figura 1, de acordo com uma concretização particular da invenção;

- as Figuras 4A e 4B mostram os dois modos de interpolação temporal de imagens capturadas de acordo com o método descrito na Figura 1, de acordo com duas concretizações da invenção;

25 - a Figura 5 mostra esquematicamente a estrutura de uma unidade de processamento de imagens, de acordo com uma concretização particular da invenção; e

- as Figuras 6 e 7 mostram um método para o processamento de imagens de vídeo implementadas em uma unidade de processamento da Figura 5, de acordo com duas concretizações particulares da invenção.

30 5. DESCRIÇÃO DETALHADA DAS CONCRETIZAÇÕES DA INVENÇÃO

A Figura 2 mostra um método para o registro temporal de imagens capturadas como descrito em relação à Figura 1, de acordo com uma concretização da invenção. Os elementos da Figura 2, que são idênticos aos elementos da Figura 1, mantêm as mesmas referências daquelas da Figura 1. Dois dispositivos de aquisição de imagens, respectivamente,
35 o esquerdo G 11 e o direito D 12, registram uma sequência de imagens de uma mesma cena de acordo com dois pontos de vista diferentes, respectivamente, um ponto de vista esquerdo e um ponto de vista direito. O resultado desses registros é mostrado pelo diagrama

de sincronização de aquisição de sequência de vídeos 13. De acordo com o exemplo de implementação da Figura 2, a sequência de vídeos esquerda compreende três imagens esquerdas sucessivas 14, 15 e 16, cada uma dessas imagens sendo representativa de uma cena compreendendo dois cilindros capturados nos tempos T1, T2 e T3, respectivamente, e a sequência de vídeos direita também compreende três imagens direitas sucessivas 17, 18 e 19, representativas da mesma cena que das imagens esquerdas, mas capturadas de acordo com outro ponto de vista do que o das imagens esquerdas, as três imagens direitas 17 a 19 sendo capturadas, respectivamente, nos tempos T1, T2 e T3. As imagens 14 e 17 são, desse modo, registradas de um modo sincronizado (isto é, ao mesmo tempo) no tempo T1, as imagens 15 e 18 são registradas de um modo sincronizado no tempo T2, e as imagens 16 e 19 são registradas de um modo sincronizado no tempo T3. Cada par de imagens esquerda e direita 14/17, 15/18 e 16/19 é adequado para formar uma imagem estereoscópica, restaurando uma vista tridimensional da cena registrada, devido à distância espacial dos dispositivos de aquisição esquerdo e direito 11 e 12, que registram a mesma cena de acordo com dois diferentes pontos de vista. Para restaurar a visão estereoscópica da cena adquirida pelos dispositivos de aquisição 11 e 12 em um dispositivo de exibição (por exemplo, um painel de exibição de plasma, uma tela LCD - de visor de cristal líquido), as imagens esquerda e direita capturadas são exibidas sequencialmente no tempo, de acordo com o seguinte esquema: imagem esquerda 14, imagem direita 17, imagem esquerda 15, imagem direita 18, imagem esquerda 16, imagem direita 19, e assim por diante. De acordo com o exemplo mostrado com relação à Figura 2, as imagens esquerdas 14, 15 e 16, respectivamente, de acordo com um método conhecido daqueles versados na técnica chamado exibição sequencial temporal. O retardo de exibição corresponde ao período de endereçamento do dispositivo de exibição 20, e é, por exemplo, 10 ms, quando a frequência de exibição é de 100 Hz. A sequência de vídeos, composta de imagens esquerdas, é vista pelo olho esquerdo de um espectador e a sequência de vídeos, composta de imagens direitas, é vista pelo olho direito de um espectador, o espectador tendo que usar óculos adaptados, que permitem a ocultação de imagens esquerdas para o olho direito e a ocultação de imagens direitas para o olho esquerdo. O desvio temporal para a exibição de imagens esquerdas e direitas, adquiridas em um modo sincronizado, leva a um defeito de exibição de imagens estereoscópicas, caracterizado por uma vibração de imagens, o defeito aparecendo quando as sequências de vídeos reproduzidas contêm elementos (por exemplo, objetos, pessoas) em movimento. Para reduzir, ou eliminar esse defeito, um registro temporal de imagens exibidas com um ligeiro retardo (ou ligeiro avanço) é conduzido. De acordo com o exemplo da Figura 2, um processo de vídeo é operado na sequência de vídeos direita. O registro temporal da sequência de vídeos, composta das imagens direitas 17, 18 e 19, é interpolado temporaria-

mente, uma sequência de vídeos interpolada, composta de imagens direitas interpoladas 17i, 18i e 19i, sendo obtida. A imagem direita interpolada 17i é obtida por interpolação temporal das imagens 17 e 18 da sequência de vídeos direita de fonte, isto é, da sequência de vídeos direita adquirida pelo dispositivo de aquisição direito 12. A imagem direita interpolada 18i é obtida por interpolação temporal das imagens 18 e 19 da sequência de vídeos direita de fonte. A imagem direita interpolada 19i é obtida por interpolação temporal da imagem 19 e da imagem diretamente seguinte à imagem 19 (não mostrada na Figura 2) da sequência de vídeos direita de fonte. Cada imagem interpolada 17i a 19i compreende uma representação de cilindros, compondo a cena adquirida. As imagens 17i, 18i e 19i compreendem todas uma representação 200 do cilindro estático e uma representação 210 do cilindro móvel em três diferentes tempos, respectivamente, $T1+\delta t$, $T2+\delta t$ e $T3+\delta t$, δt correspondendo ao desvio temporal entre a imagem-fonte e a imagem interpolada correspondente, isto é, o desvio temporal entre o tempo de aquisição da imagem-fonte na sequência de vídeos de fonte e o tempo de exibição da imagem interpolada na sequência de vídeos exibida correspondente. Para reduzir ou eliminar os defeitos de exibição ligados à exibição de imagens-fonte, a interpolação de imagens direitas é conduzida com compensação de movimento. A interpolação com compensação de movimento permite que a posição dos cilindros 200 e 201 seja determinada no momento de exibição da imagem interpolada, isto é, a $T1+\delta t$ para a imagem 17i, a $T2+\delta t$ para a imagem 18i e a $T3+\delta t$ para a imagem 19i. A interpolação da imagem de movimento compensado, aplicada à sequência de vídeos direita, propicia, desse modo, que as imagens interpoladas, que vão ser geradas, tenham um conteúdo correspondente ao tempo de captura desviado de δt , isto é, desviado da duração que há entre a exibição da imagem esquerda e de uma imagem direita no dispositivo de exibição, que não corresponda ao tempo de captura real da imagem-fonte associada (isto é, da imagem-fonte 17 para a imagem interpolada 17i, da imagem-fonte 18 para a imagem interpolada 18i, e da imagem-fonte 19 para a imagem interpolada 19i).

De acordo com uma variante, as imagens esquerdas 14 a 16 são exibidas com um ligeiro avanço com relação às imagens direitas 17 a 19, respectivamente. De acordo com essa variante, a interpolação de movimento compensado é conduzida na sequência de vídeos esquerda compreendendo as imagens esquerdas.

De acordo com uma outra variante, as imagens esquerdas 14 a 16 são exibidas com um ligeiro retardo com relação às imagens direitas 17 a 19, respectivamente. De acordo com essa variante, a interpolação com compensação de movimento é conduzida na sequência de vídeos esquerda, compreendendo as imagens esquerdas.

De acordo com outra variante, a interpolação com compensação de movimento é conduzida na sequência de vídeos esquerda e na sequência de vídeos direita.

De acordo com uma variante, os pares de imagens 14 e 17, de imagens 15 e 18 e

de imagens 16 a 19 são registrados de um modo sincronizado, em comparação a um relógio comum, isto é, as imagens 14 e 17 são capazes de serem adquiridas com um determinado retardo, as imagens 15 e 18 são capazes de serem adquiridas com o mesmo determinado retardo e as imagens 16 e 19 são capazes de serem adquiridas com o mesmo determinado

5 retardo.

A Figura 3 mostra um método para interpolação com compensação de movimento de imagens-fonte, capturadas de acordo com o método descrito com relação às Figuras 1 e 2, de acordo com uma concretização não restritiva da invenção. A imagem I 31, a ser interpolada, é interpolada de duas imagens-fonte 30 e 32 de uma sequência de vídeos de fonte,

10 a primeira 30 dessas imagens-fonte sendo chamadas de imagem-fonte precedente T-1, e a segunda 32 dessas imagens sendo chamada de imagem-fonte corrente T. O desvio temporal entre a imagem-fonte precedente T-1 30 e a imagem a ser interpolada I 31 é igual a δt_2 . A interpolação com compensação de movimento compreende, vantajosamente, uma etapa de estimativa de movimento seguida por uma etapa de interpolação. A estimativa de movimento

15 consiste em definir um vetor de movimento para cada um dos pontos da imagem a ser interpolada I 31, a dita imagem sendo posicionada temporariamente entre as duas imagens-fonte T-1 30 e T 32. A etapa de estimativa de movimento é conduzida vantajosamente em 2 etapas: uma etapa de previsão e uma etapa de correção. A etapa de previsão consiste em definir, para cada um dos pixels da imagem a ser interpolada I 31, um vetor de movimento,

20 dos vetores de movimento já calculados para os pixels vizinhos, e da projeção de um vetor de movimento, calculada para a imagem-fonte precedente T-1 30, e passando pelo pixel considerado da imagem a ser interpolada I 31. Os vetores de movimento calculados são, por exemplo, aqueles dos n pixels vizinhos da linha de pixels precedente. A previsão consiste então em selecionar, entre os n+1 vetores de movimento pré-calculado, o vetor de movimento gerando o menor valor DFD (Diferença de Quadro de Deslocamento). A etapa

25 de correção consiste então em girar em torno do seu eixo, o vetor de movimento selecionado para reduzir ainda mais, se possível, a DFD do pixel considerado da imagem a ser interpolada I 31. A estimativa de movimento é conduzida vantajosamente para vários pixels da imagem a ser interpolada I 31, por exemplo, 25% da totalidade dos pixels, ou um pixel em

30 quatro ou 50% da totalidade de pixels, ou um pixel em dois. De acordo com uma variante, a estimativa de movimento é conduzida para o conjunto de pixels da imagem a ser interpolada I 31. De acordo com uma outra variante, a etapa de estimativa de movimento é conduzida em uma única etapa.

A etapa de interpolação, que se segue à etapa de estimativa de movimento, é, vantajosamente, uma interpolação bilinear. No caso a imagem a ser interpolada I 31, entre uma

35 imagem-fonte precedente T-1 30 e uma imagem-fonte corrente T 32, o vetor de movimento, definido previamente, aponta para um ponto de imagem de cada uma das imagens-fonte,

esse ponto de imagem sendo um pixel, ou um ponto situado entre 4 pixels. No caso de uma interpolação de quadro único, a interpolação bilinear consiste então em atribuir ao pixel considerado P_i 301 da imagem a ser interpolada um valor, que é uma função bilinear f de valores de 4 pixels apontados a uma das imagens-fonte (por exemplo, a imagem-fonte corrente) vetor de movimento estimado para o pixel P_i . Por exemplo, se NV_1, NV_2, NV_3 e NV_4 indicam os valores (por exemplo, o nível de vídeo NV deles) de 4 pixels próximos ao ponto apontado para a imagem P_T 302 pelo vetor de movimento, na imagem-fonte corrente T32, e se $\alpha_1, \alpha_2, \alpha_3$ e α_4 são fatores de ponderação, representativos da proximidade dos ditos pixels com o ponto de imagem P_T , o valor atribuído ao pixel P_i 301 da imagem a ser interpolada I 31 é

10
$$\left(\sum_{i=1}^4 \alpha_i \cdot NV_i \right) / \sum_{i=1}^4 \alpha_i$$
. Se um vetor de movimento aponta a um pixel particular, entre os 4 pixels, o fator de ponderação, atribuído aos 3 outros pixels é nulo, e o valor atribuído ao pixel considerado é o valor desse pixel particular. No caso de uma interpolação de quadro duplo, a interpolação bilinear é feita do mesmo modo, mas com 8 pixels, isto é, 4 pixels da imagem-fonte precedente e 4 pixels da imagem-fonte corrente. De acordo com uma variante, a imagem-fonte apontada pelo vetor de movimento (ou o seu oposto) é a imagem-fonte precedente T-1 30, e o ponto de imagem apontado é P_{T-1} 300. De acordo com uma variante, a interpolação é conduzida, por exemplo, por interpolação linear temporal de uma área fixa, por filtragem espacial e temporal nas áreas críticas da imagem (isto é, as áreas para as quais a confiança do vetor de movimento é baixa), ou por qualquer método conhecido daqueles versados na técnica. De acordo com uma outra variante, a interpolação é conduzida por qualquer combinação dos métodos citados previamente.

De acordo com uma variante, a interpolação de movimento compensado é conduzida em uma única etapa.

As Figuras 4A e 4B mostram dois modos de interpolação temporal de imagens-fonte, capturadas de acordo com o método descrito em relação às Figuras 1 e 2, de acordo com duas concretizações particulares da invenção.

A Figura 4A mostra um primeiro exemplo de interpolação temporal de movimento compensado. Na Figura 4A, os picos 401g, 402g e 403 g, em linhas sólidas, representam os imagens esquerdas de fonte adquiridas por um dispositivo de aquisição esquerdo. A imagem esquerda 402g corresponde a uma imagem chamada a imagem corrente T, a imagem esquerda 401g tem uma imagem chamada imagem precedente T-1, e a imagem esquerda 403g tem uma imagem chamada a imagem seguinte T+1, as três imagens esquerdas 401g a 403g pertencentes à sequência de vídeos esquerda e sendo sucessivas de um ponto de vista temporal. Os picos 401d, 402d e 403d, em linhas sólidas, representam as imagens direitas de fonte por um dispositivo de aquisição direito. A imagem direita 402d corresponde a uma imagem chamada a imagem corrente T, a imagem direita 401d tem uma imagem

chamada imagem precedente T-1, e a imagem direita 403d tem uma imagem chamada a imagem seguinte T+1, as três imagens direitas 401d a 403d pertencentes à sequência de vídeos direita e sendo sucessivas de um ponto de vista temporal. Como mostrado na Figura 4A, a imagem esquerda 401g foi adquirida em um modo sincronizado temporalmente com a

5 imagem direita 401d, a imagem esquerda 402g foi adquirida em um modo sincronizado temporalmente com a imagem direita 402d, e a imagem esquerda 403g foi adquirida em um modo sincronizado temporalmente com a imagem direita 403d. Os picos 404i e 405i, em linhas pontilhadas, representam imagens direitas interpoladas de imagens direitas adquiridas pelo dispositivo de aquisição, também chamadas imagens direitas de fonte. A imagem

10 direita interpolada 404i é interpolada das imagens direitas de fonte 401d e 402d, e a imagem direita interpolada 405i é interpolada das imagens direitas de fonte 402d e 403d. De acordo com o exemplo da Figura 4A, a imagem direita interpolada 404i é situada no ponto intermediário de um ponto de vista temporal, das imagens direitas de fonte 401d e 402d, e a imagem direita interpolada 405i é situada no ponto intermediário de um ponto de vista temporal,

15 das imagens direitas de fonte 402d e 403d. Se a duração separando duas imagens-fonte, por exemplo, 401d e 402d, corresponder a um tempo de quadro T, a duração δt_1 , separando a imagem-fonte 401d da imagem interpolada 404i, é igual à duração δt_2 , separando a imagem-fonte 402d da imagem interpolada 404i, e é igual a 1/2 tempo de quadro. Do mesmo modo, a imagem interpolada seguinte 405i é separada temporalmente de ambos os quadros

20 direito de fonte 402d e 403d, circundando o 1/2 tempo de quadro. Essa implementação exemplificativa oferece a vantagem de conduzir a interpolação temporal com compensação de movimento apenas em uma das duas sequência de vídeos de fonte, isto é, a sequência de vídeos de fonte direita. De acordo com uma variante, a interpolação com compensação de movimento é conduzida na sequência de vídeos de fonte esquerda G e não na sequência

25 de vídeos de fonte direita D.

A Figura 4B mostra um segundo exemplo de interpolação temporal com compensação de movimento. Na Figura 4B, os picos 401g, 402g, 403g, 401d, 402d e 403d, em linhas sólidas, representam as imagens-fonte esquerda e direita, e são idênticos àqueles descritos com relação à Figura 4A, que tem as mesmas referências. Os picos 406i e 408i, em linhas

30 pontilhadas, representam as imagens esquerdas interpoladas, a imagem interpolada 406i sendo interpolada das imagens-fonte esquerdas 401g e 402g, e a imagem interpolada 408i sendo interpolada as imagens-fonte esquerdas 402g e 403g. De acordo esse exemplo, a duração separando a imagem-fonte esquerda 401g da imagem interpolada 406i é igual a 3/4 do tempo de quadro, e a duração separando a imagem interpolada 406i da imagem-fonte

35 esquerda corrente 402g é igual a 1/4 do tempo de quadro. A imagem interpolada 406i é temporalmente mais próxima da imagem-fonte esquerda corrente 402g do que a imagem-fonte esquerda precedente 401g, isto é, três vezes temporalmente mais próxima. Do mesmo

modo, a duração separando a imagem-fonte esquerda 402g da imagem interpolada 408i é igual a $\frac{3}{4}$ do tempo de quadro, e a duração separando a imagem interpolada 408i da imagem-fonte esquerda seguinte 403g é igual a $\frac{1}{4}$ do tempo de quadro. A imagem interpolada 408i é temporalmente mais próxima da imagem-fonte esquerda seguinte 403g do que a

5 imagem-fonte esquerda corrente 402g, isto é, três vezes temporalmente mais próxima. Os picos 406i e 408i, em linhas pontilhadas, representam as imagens esquerdas interpoladas, a imagem interpolada 406i sendo interpolada das imagens-fonte esquerdas 401g e 402g, e a imagem interpolada 408i sendo interpolada das imagens-fonte esquerdas 402g e 403g. De acordo com esse exemplo, a duração separando a imagem-fonte direita precedente 401d da

10 imagem interpolada 407i é igual a $\frac{1}{4}$ do tempo de quadro, e a duração separando a imagem interpolada 407i da imagem-fonte direita corrente 402d é igual a $\frac{3}{4}$ do tempo de quadro. A imagem interpolada 407i é temporalmente mais próxima da imagem-fonte direita precedente 401d do que a imagem-fonte direita precedente 402d, isto é, três vezes temporalmente mais próxima. Do mesmo modo, a duração separando a imagem-fonte direita corrente

15 te 402gd da imagem interpolada 409i é igual a $\frac{1}{4}$ do tempo de quadro, e a duração separando a imagem interpolada 409i da imagem-fonte direita seguinte 403d é igual a $\frac{3}{4}$ do tempo de quadro. A imagem interpolada 409i é temporalmente mais próxima da imagem-fonte direita corrente 402d do que a imagem-fonte direita seguinte 403d, isto é, três vezes temporalmente mais próxima. Se, por exemplo, mostrada na Figura 4A, ou, por exemplo,

20 mostrada na Figura 4B, a diferença temporal no nível de exibição entre uma imagem esquerda (fonte para a Figura 4A, interpolada para a Figura 4B), e uma imagem direita (interpolada para as Figuras 4A e 4B) é constante e igual a $\frac{1}{2}$. A implementação exemplificativa oferece a vantagem de minimizar os erros ligados à interpolação temporal com compensação de movimento, a imagem interpolada sendo temporalmente próxima a uma imagem-

25 fonte (isto é, $\frac{1}{4}$ de quadro).

De acordo com uma variante, várias imagens interpoladas são geradas entre duas imagens-fonte, por exemplo, 2, 3 ou 4 imagens interpoladas. A geração de várias imagens interpoladas, entre duas imagens-fonte, permite que várias imagens para exibição sejam multiplicadas, e, conseqüentemente, aumentem a frequência de exibição de imagens no

30 dispositivo de exibição. De acordo com essa variante, as imagens-fonte e as imagens interpoladas são vantajosamente distribuídas uniformemente de um ponto de vista temporal, isto é, a duração que as separa é idêntica para todas as imagens.

A Figura 5 mostra esquematicamente uma concretização de hardware de uma unidade de processamento 5 de um dispositivo de exibição, de acordo com uma concretização particular da invenção. A unidade de processamento 5 tem, por exemplo, a forma de um

35 circuito lógico programável do tipo FPGA (Arranjo de Circuitos de Campo Programável), por exemplo, ASIC (Circuito Integrado de Aplicação Específica) ou um DSP (Processador de

Sinais Digitais). A unidade de processamento é adaptada para a implementação do método descrito previamente.

A unidade de processamento 5 compreende os seguintes elementos:

- um pré-filtro 50;
- 5 - um estimador de movimento 51;
- memórias 52, 53 e 56, por exemplo, do tipo RAM (Memória de Acesso Aleatório) ou instantânea, capazes de memorizar uma ou mais várias imagens-fonte;
- um interpolador 54;
- uma unidade de sincronização 56;
- 10 - barramentos de dados 510, 511, 512, 513, 514 e 515; e
- um barramento de relógio 516.

Um primeiro sinal RVBd (para imagem direita Vermelha, Verde, Azul) 510, representativo da sequência de vídeos de fonte direita, compreendendo imagens-fonte direitas adquiridas pelo dispositivo de aquisição direito, é suprido à entrada da unidade de processamento 5, a um pré-filtro 50, por meio de um barramento de dados 511. O pré-filtro 50 filtra os pixels compondo as imagens-fonte esquerdas, isto é, uma imagem-fonte corrente T e uma imagem-fonte precedente T-1, enquanto determinando por cálculo, por exemplo, um pixel para quatro pixels da imagem-fonte, ou um pixel para dois pixels da imagem-fonte, de acordo com um esquema espacial determinado, para armazenar vários pixels de todas as imagens-fonte corrente T e as imagens precedentes T-1 na memória 52. O estimador de movimento 51 determina um vetor de movimento (V_x , V_y) para cada pixel da imagem a ser interpolada, todos os pixels, para os quais um vetor de movimento é determinado, sendo associados com vários pixels (4 ou 2 de acordo com os exemplos apresentados acima) de imagens-fonte correntes T e de imagens-fonte precedentes T-1, armazenadas na memória 52. Um mesmo vetor de movimento (V_x , V_y) é associado com vários pixels de imagens-fonte T e T-1. Esses vetores de movimento determinados (V_x , V_y) são transmitidos para o interpolador 513 pelo barramento de dados 513. Em paralelo, o interpolador 54 recebe, na entrada, o sinal RVBd 501, representativo de imagens-fonte direitas correntes T e imagens-fonte direitas precedentes T-1, os níveis de vídeo de todos os pixels, de todas as imagens correntes T e das imagens precedentes T-1, sendo armazenados na memória 53. O interpolador 54 gera uma imagem direita interpolada, de imagens-fonte direitas correntes T e da imagem-fonte direita precedente T-1, e vetores de movimento (V_x , V_y) associados. O sinal RVBd,i (para imagem direita Vermelha, Verde, Azul interpolada), representativo de imagens direitas interpoladas da sequência de vídeos direita de fonte, é transmitido a uma unidade de exibição 502, vantajosamente, externa à unidade de processamento 5, pelo barramento de dados 515, para exibir uma sequência de vídeos direita interpolada, compreendendo as imagens direitas interpoladas. A sequência de vídeos direita interpolada é composta de ima-

gens direitas interpoladas em movimento, que substituem as imagens-fonte direitas compreendidas na sequência de vídeos direita de fonte. Desse modo, as imagens direitas exibidas no dispositivo de exibição são imagens direitas interpoladas e não imagens-fonte direitas. Com um processamento de vídeo sendo operado na sequência de vídeos direita de fonte,

5 um retardo é gerados no nível de transmissão de imagens interpoladas no dispositivo de exibição 502, com relação às imagens esquerdas associadas, para as quais o conteúdo corresponde a todas as imagens direitas interpoladas. Para sincronizar a transmissão do sinal RVBg (para a imagem esquerda Vermelha, Verde, Azul) 500 na unidade de exibição 502 com o sinal RVBd,i 504, o sinal RVBg 500 é retardado por memorização de uma ou mais

10 imagens esquerdas em uma memória 56, pela parte intermediária de uma unidade de sincronização 55, o sinal RVBg 500 sendo suprido à unidade de sincronização 55 pelo barramento de dados 510. A unidade de sincronização 55 também transmite um sinal de relógio 54 para o interpolador 54, pelo barramento de relógio 516, para disparar a transmissão de uma imagem direita interpolada para a unidade de exibição 502, de tal maneira que a ima-

15 gem direita interpolada é recebida pela unidade de exibição, logo após a imagem esquerda, para a qual o conteúdo corresponde à imagem direita interpolada, para gerar uma imagem estereoscópica de um par de imagens, compreendendo uma imagem esquerda e uma imagem direita interpoladas associadas. A imagem esquerda do par de imagens, intencionada para formar uma imagem estereoscópica, é transmitida pela unidade de sincronização 55 a

20 uma unidade de exibição 502, por meio de um barramento de dados 514.

De acordo com uma variante, a unidade de processamento 5 não compreende a decimalização de imagens-fonte, isto é, que a estimativa de movimento é conduzida para todos os pixels compondo as imagens-fonte precedentes T-1 e as imagens-fonte correntes T. De acordo com essa variante, o pré-filtro é funcionalmente menos complexo (apenas

25 conduzindo, por exemplo, nivelamento gaussiano de imagens-fonte) do que quando um pixel é determinado por vários pixels de imagens-fonte. De acordo com essa variante, a memória 52 apenas armazena uma única imagem, isto é, a imagem precedente, a imagem corrente sendo transmitida diretamente para o estimador de movimento, sem ser memorizada de antemão. Do mesmo modo, a memória 53 apenas armazena uma única imagem, isto

30 é, a imagem precedente, a interpolação sendo conduzida da imagem corrente, recebida pelo interpolador, e da imagem precedente memorizada. Após interpolação, a imagem corrente se torna a imagem precedente de uma nova imagem corrente e é armazenada na memória 53, sobrepondo-se à imagem precedente velha.

De acordo com uma outra variante, as memórias 52 e 53 formam apenas uma única memória.

35

De acordo com uma outra variante, o sinal RVBg, representativo da sequência de vídeos esquerda de fonte, é também interpolado com compensação de movimento. De

acordo com essa variante, a unidade de processamento 5 compreende um segundo pré-filtro, um segundo estimador de movimento, um segundo interpolador com as memórias associadas para a interpolação com compensação de movimento de imagens esquerdas, além do pré-filtro 50, o estimador de movimento 51, o interpolador 54 e as suas memórias associadas 52 e 53, intencionadas para interpolação de imagens direitas. O funcionamento dos pré-filtros secundários, estimadores de movimento e interpoladores é idêntico ao descrito previamente para o processamento de imagens direitas.

A Figura 6 mostra um método para processamento de imagens de vídeo, implementado em uma unidade de processamento 5, de acordo com uma concretização particularmente vantajosa não restritiva da invenção.

Durante uma etapa de inicialização 60, os diferentes parâmetros da unidade de processamento são atualizados.

Depois, durante uma etapa 61, uma imagem é interpolada de duas imagens-fonte de uma sequência de vídeos, a sequência de vídeos resultando da combinação de suas sequências de vídeos, chamadas a sequência de vídeos esquerda e a sequência de vídeos direita, as superfícies de vídeo esquerda e direita compreendendo, respectivamente, as imagens esquerdas e as imagens direitas de uma mesma cena filmada, respectivamente, por uma câmera esquerda e uma câmera direita, em um modo sincronizado. A sequência de vídeos, resultante da combinação de sequências de vídeo esquerda e direita, é composta de imagens esquerdas e direita de sequências de vídeo esquerdas e direitas, sucedendo-se, temporalmente, entre si, alternadamente, isto é, as imagens esquerdas e direitas são temporalmente alternadas da seguinte maneira: imagem esquerda, imagem direita, imagem esquerda, imagem direita, imagem esquerda e assim por diante. Em outras palavras, a sequência de vídeos resultante da combinação de sequências de vídeo esquerdas e direitas compreende uma pluralidade de pares de imagens esquerdas e direitas, que se sucedem entre si temporalmente, cada par de imagens compreendendo uma imagem esquerda e uma imagem direita, representativas da mesma cena, capturadas no mesmo tempo T . Cada par de imagens é intencionado para gerar uma imagem estereoscópica, isto é, uma imagem em três dimensões, da cena. A imagem esquerda de um par de imagens é também chamada a primeira imagem, a imagem direita de um par de imagens é também chamada a segunda imagem, e a sequência de vídeos, resultante da combinação de sequências de vídeo esquerda e direita, é chamada a primeira sequência de vídeos, no restante da descrição. A imagem interpolada, também chamada a terceira imagem, é interpolada, vantajosamente, com compensação de movimento de duas imagens secundárias sucedendo-se temporalmente na primeira sequência de vídeos. A terceira imagem é interposta temporalmente entre as duas imagens secundárias. O intervalo temporal, separando a segunda imagem, precedendo (de um ponto de vista temporal) a terceira imagem, da terceira imagem, é igual a δt_1 ,

e o intervalo temporal separando a segunda imagem, seguinte (de um ponto de vista temporal) à terceira imagem, da terceira imagem é igual a δt_2 . Vantajosamente, δt_1 é igual a δt_2 .

De acordo com uma variante, várias imagens terciárias são interpoladas com compensação de movimento das imagens secundárias, acentuadamente, para aumentar a frequência de imagens da primeira sequência de vídeos.

De acordo com uma outra variante, a terceira imagem ou as imagens terciárias é ou são interpostas de duas imagens esquerdas seguindo-se temporalmente entre si. De acordo com essa variante, a imagem esquerda de um par de imagens é chamada a segunda imagem, e a imagem direita de um par de imagens é chamada a primeira imagem, para manter o mesmo vocabulário apresentado acima.

De acordo com uma outra variante, a terceira imagem é interpolada com compensação de movimento de duas imagens direitas, chamadas imagens secundárias, e uma quarta imagem resulta da interpolação com compensação de movimento de duas imagens esquerdas, chamadas imagens primárias. De acordo com essa variante, o intervalo temporal separando a primeira imagem precedente (de um ponto de vista temporal) da quarta imagem, da quarta imagem é igual a δt_1 , e o intervalo temporal separando a primeira imagem seguinte (de um ponto de vista temporal) à quarta imagem, da quarta imagem é igual a δt_2 . De acordo com essa variante, δt_1 é vantajosamente estritamente inferior a δt_2 , isto é, que a terceira imagem é temporalmente mais próxima do que a segunda imagem que a precede um $\delta t_1'$, que é estritamente superior a $\delta t_2'$, isto é, que a quarta imagem é temporalmente mais próxima da primeira imagem, que a segue, e o intervalo temporal separando a terceira imagem da quarta imagem é igual à metade do tempo separando as duas imagens primárias, na sequência de vídeos esquerda (isto é, aquelas capturadas por um dispositivo de aquisição).

A interpolação com compensação de movimento da terceira imagem e/ou da quarta imagem é vantajosamente reiterada, de modo a substituir todas as imagens secundárias da primeira sequência de vídeos com as imagens interpoladas, isto é, as imagens terciárias, e/ou substituir todas as imagens primárias da primeira sequência de vídeos com as imagens interpoladas, isto é, as imagens quaternárias.

A Figura 7 mostra um método para o processamento de imagens de vídeo implementadas em uma unidade de processamento 5, de acordo com uma concretização particularmente vantajosa não restritiva da invenção.

Durante uma etapa de inicialização 70, os diferentes parâmetros da unidade de processamento são atualizados.

Depois, durante uma etapa 71, um vetor de movimento é calculado para cada pixel de vários pixels da imagem a ser interpolada, isto é, a terceira imagem e/ou a quarta imagem, por exemplo, para um pixel em dois ou para um pixel em quatro. De acordo com uma

variante, um vetor de movimento é determinado para cada pixel da invenção. Para a terceira imagem, os vetores de movimento são determinados de duas imagens secundárias, e para a quarta imagem, os vetores de movimento são determinados de duas imagens primárias.

Depois, durante uma etapa 72, a terceira imagem e/ou a quarta imagem é ou são
5 interpostas das duas imagens secundárias (respectivamente, de duas imagens primárias) e de vetores de movimento determinadas na etapa 71. Quando um vetor de movimento tiver sido determinado para apenas alguns pixels da terceira imagem, por exemplo, para um pixel em quarto, o vetor de movimento, calculado para o pixel entre os quatro, é copiado para o outros três pixels, de um modo a conduzir a interpolação para o conjunto de pixels da ima-
10 gem a ser interpolada, isto é, da terceira imagem. Para o restante, a etapa 72 é idêntica à etapa 61, descrita com relação à Figura 6.

Finalmente, durante uma etapa 73, as primeira e terceira imagens são exibidas em um dispositivo de exibição em uma maneira sequencial temporal, isto é, sucessiva e alternadamente, como uma primeira imagem seguida por uma terceira imagem e assim por dian-
15 te. A exibição de imagens terciárias interpoladas com compensação de movimento, em vez de imagens secundárias originais, acarreta que o defeito, ligado a uma exibição de deslocamento de uma segunda imagem com relação à primeira imagem, que é associado com ela (isto é, capturado no mesmo instante T), seja corrigido, a exibição da segunda imagem seguinte à exibição da primeira imagem. De fato, a terceira imagem substituindo a segundo
20 imagem, tendo sido registrada temporalmente (por interpolação temporal com compensação de movimento), o posicionamento temporal da terceira imagem, na sequência de vídeos exibida no dispositivo de exibição, é coerente com o seu conteúdo de movimento compensado. De acordo com uma variante, as imagens quaternárias e as imagens terciárias são exibidas em um dispositivo de exibição, de uma maneira sequencial temporal, isto é, suces-
25 siva e alternadamente, como uma quarta imagem seguida por uma terceira imagem e assim por diante.

Naturalmente, a invenção não é limitada às concretizações descritas previamente.

Em particular, a invenção não é limitada a um método para processar imagem, mas se estende à unidade de processamento, que implementa esse método, e ao dispositivo de
30 exibição, que compreende uma unidade de processamento que implementa o método de processamento de imagens.

Vantajosamente, cada uma das sequências de vídeo esquerdas e direitas compreende mais de três imagens. O número de imagens depende da duração das sequências de vídeo (por exemplo, adquiridas a 50 ou 60 Hz em modo de vídeo, isto é, quer dizer 50 ou 60
35 imagens por segundo, ou a 24 Hz em um modo filme, isto é, quer dizer 24 imagens por segundo).

REIVINDICAÇÕES

1. Método para processar vídeo de uma sequência de pares de imagens estereoscópicas, cada par de imagens estereoscópicas compreendendo uma primeira imagem (14, 15, 16) e uma segunda imagem (17, 18, 19), adquiridas em um modo sincronizado, cada
5 primeira imagem sendo uma imagem à esquerda e cada segunda imagem sendo uma imagem à direita, **CARACTERIZADO** pelo fato de que o dito método compreende:

gerar (61, 72) pelo menos uma terceira imagem (17i, 18i, 19i) por interpolação temporal de movimento compensado de pelo menos duas imagens secundárias (17, 18, 19), a pelo menos uma terceira imagem (407i) sendo gerada a partir de duas imagens secundárias (401d, 402d) imediatamente sucessivas a partir de um ponto de vista temporal, um primeiro intervalo temporal separando a terceira imagem (407i) da segunda imagem (401d) precedendo temporalmente a terceira imagem (407i), um segundo intervalo temporal separando a terceira imagem (407i) da segunda imagem (402d) temporalmente seguinte à terceira imagem (407i), o primeiro intervalo temporal sendo estritamente inferior ao segundo intervalo temporal,
15

gerar pelo menos uma quarta imagem (406i) por interpolação temporal de movimento compensado de pelo menos duas imagens primárias, a pelo menos uma quarta imagem sendo gerada a partir de duas imagens primárias (401g, 402g) imediatamente sucessivas a partir de um ponto de vista temporal, um terceiro intervalo temporal separando a quarta imagem (407i) da primeira imagem (401g) precedendo temporalmente a quarta imagem (406i), um quarto intervalo temporal separando a quarta imagem (406i) da primeira imagem (402g) temporalmente seguinte à quarta imagem (406i), o quarto intervalo temporal sendo estritamente inferior ao terceiro intervalo temporal;
20

a segunda imagem precedendo temporalmente a terceira imagem e a primeira imagem precedendo temporalmente a quarta imagem formando um primeiro par de imagens estereoscópicas da dita sequência, a segunda imagem temporalmente seguinte à terceira imagem e a primeira imagem temporalmente seguinte à quarta imagem formando um segundo par de imagens estereoscópicas da dita sequência.
25

2. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que compreende adicionalmente uma etapa de estimativa (71) de um vetor de movimento para cada pixel de uma pluralidade de pixels da dita terceira imagem (17i, 18i, 19i), a partir das ditas pelo menos duas imagens secundárias.
30

3. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que compreende adicionalmente uma etapa de exibição (73) das ditas primeira e terceira imagens em um dispositivo de exibição, as ditas primeira imagem (14, 15, 16) e terceira imagem (17i, 18i, 19i) sendo exibidas, respectivamente, sequencialmente.
35

4. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que

compreende adicionalmente uma etapa de exibição das ditas terceira e quarta imagens em um dispositivo de exibição, as ditas terceira e quarta imagens sendo, respectivamente, exibidas sequencialmente.

5 5. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o primeiro intervalo temporal e o quarto intervalo temporal são iguais, e em que o segundo intervalo temporal e o terceiro intervalo temporal são iguais.

10 6. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que uma pluralidade de terceiras imagens é gerada a partir de duas segundas imagens (401d, 402d) imediatamente sucessivas a partir de um ponto de vista temporal, e em que uma mesma pluralidade de quartas imagens é gerada a partir de duas primeiras imagens (401d, 402d) imediatamente sucessivas a partir de um ponto de vista temporal.

15 7. Método, de acordo com a reivindicação 6, **CARACTERIZADO** pelo fato de que as terceiras imagens, que formam a dita pluralidade, são distribuídas uniformemente a partir de um ponto de vista temporal entre as ditas duas segundas imagens (401d, 402d), e em que as quartas imagens, que formam a dita pluralidade, são distribuídas temporalmente entre as ditas duas primeiras imagens (401d, 402g), do mesmo modo que as terceiras imagens.

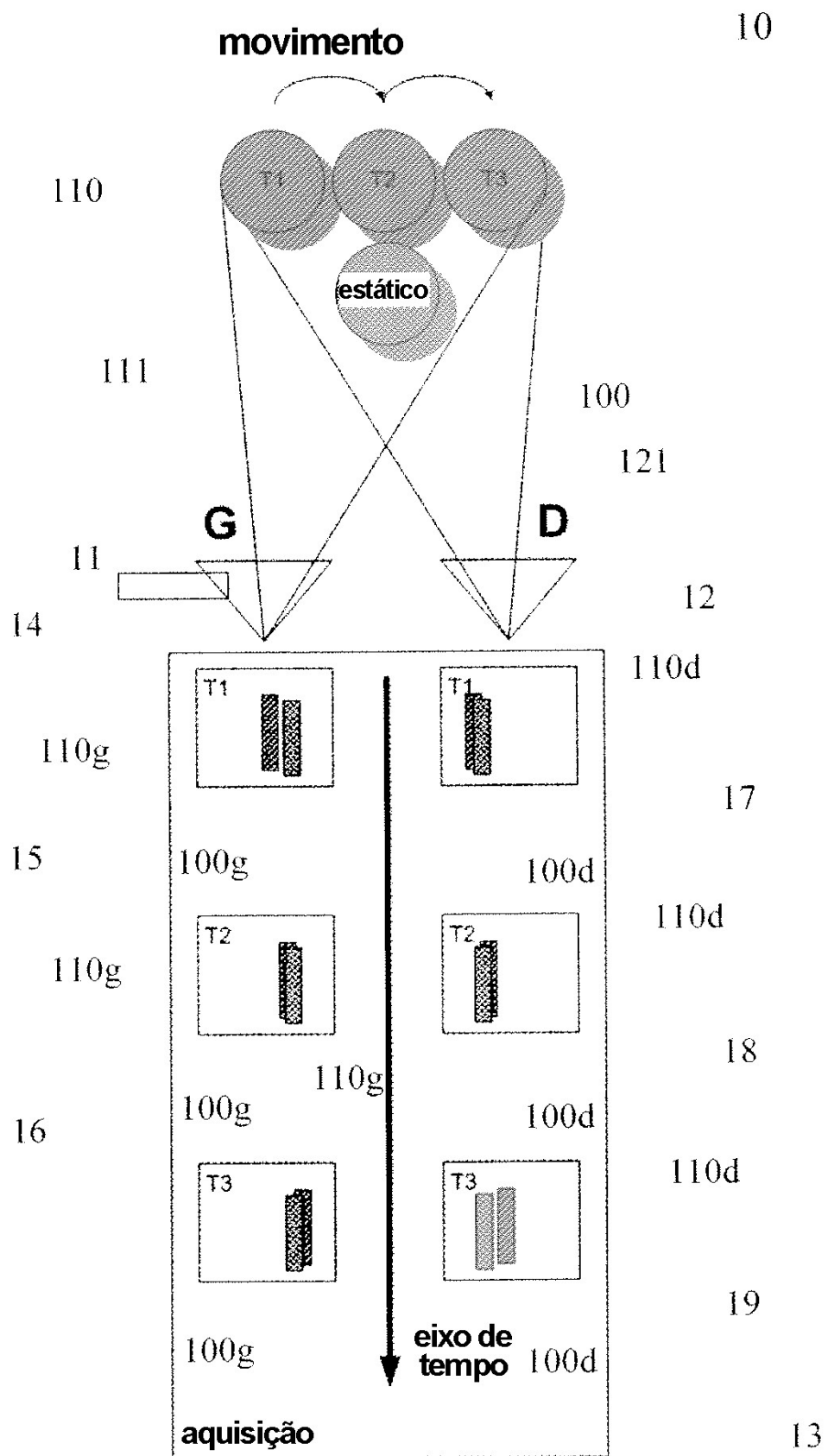


Fig 1

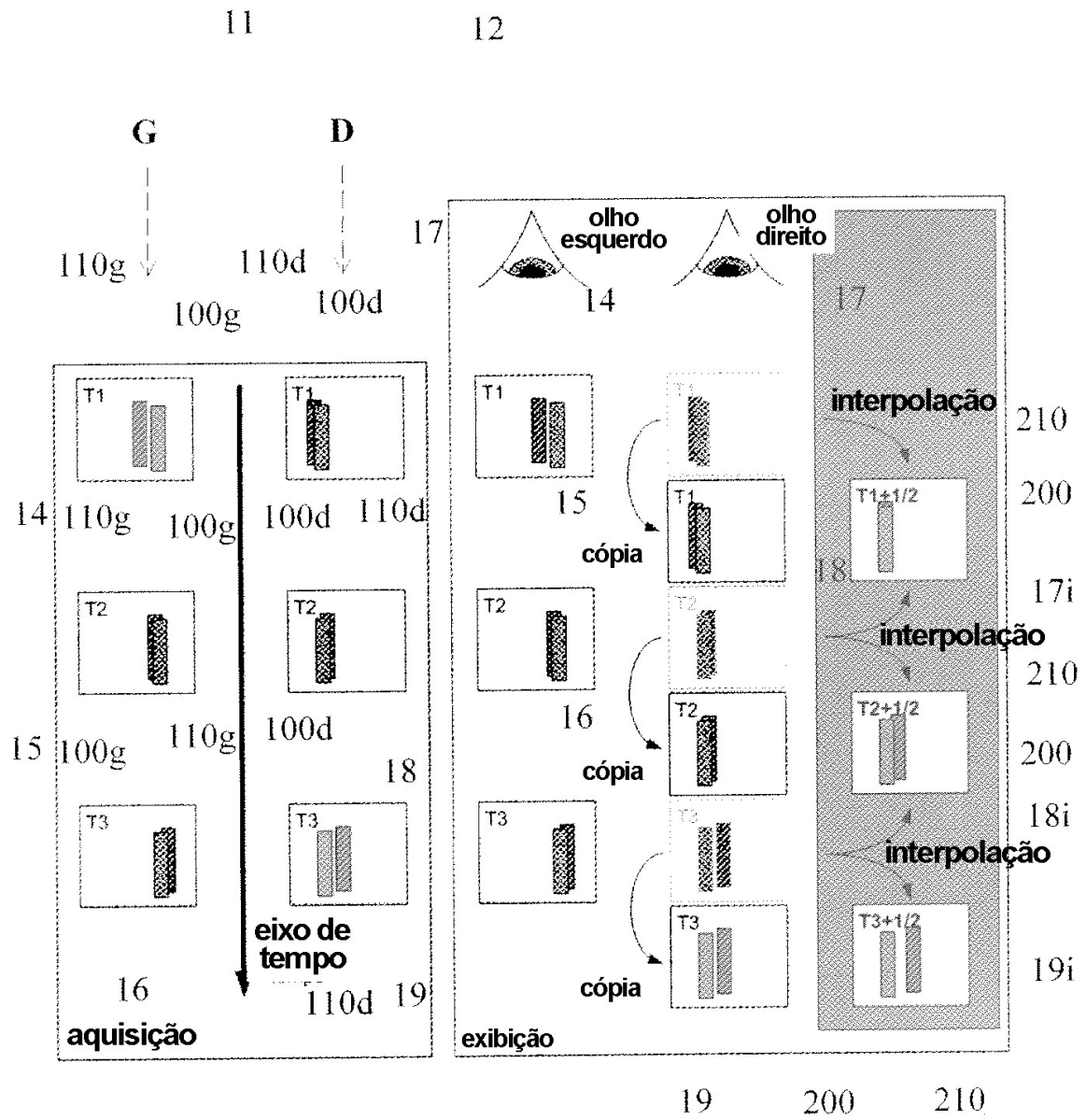


Fig 2

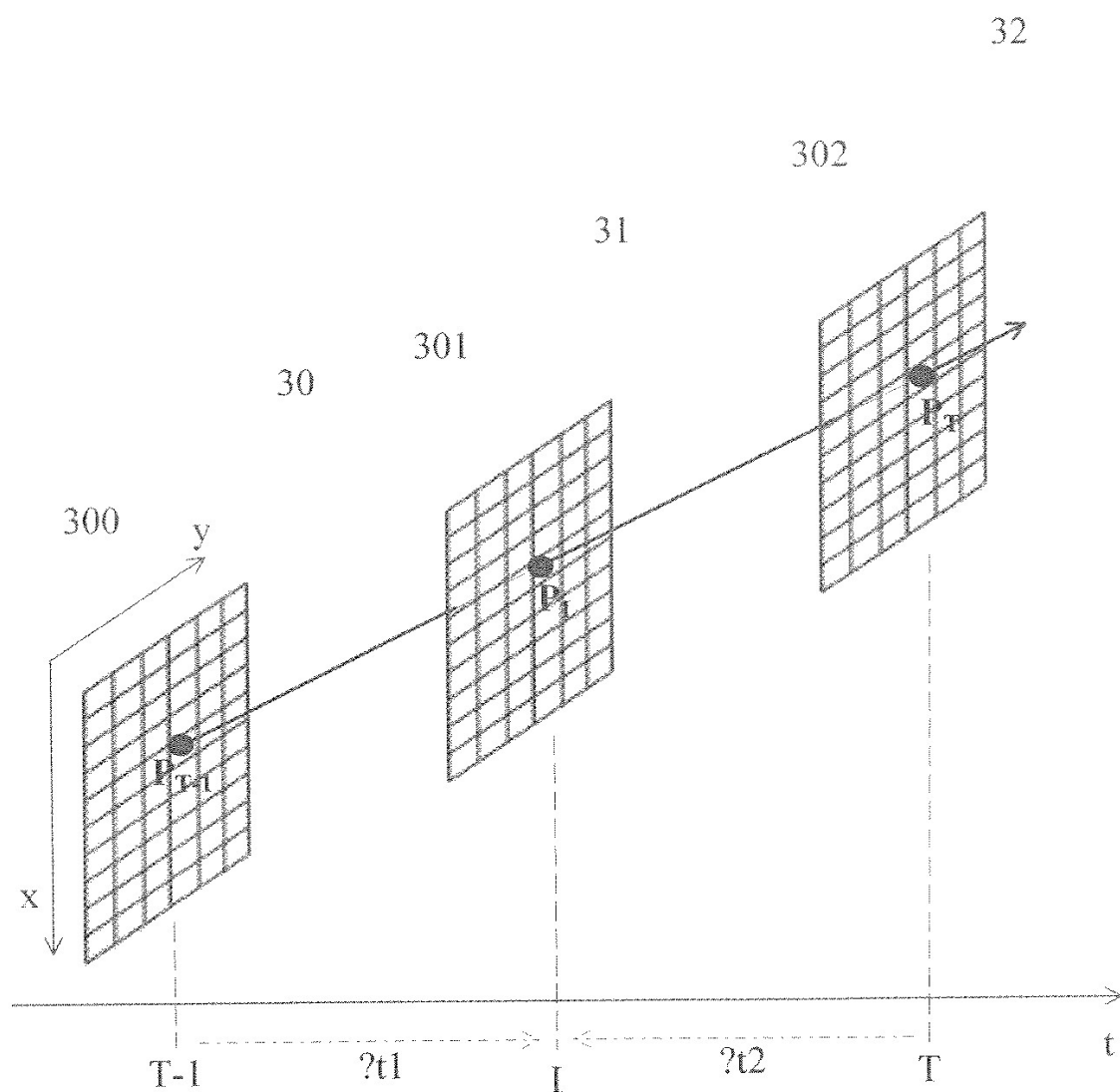


Fig 3

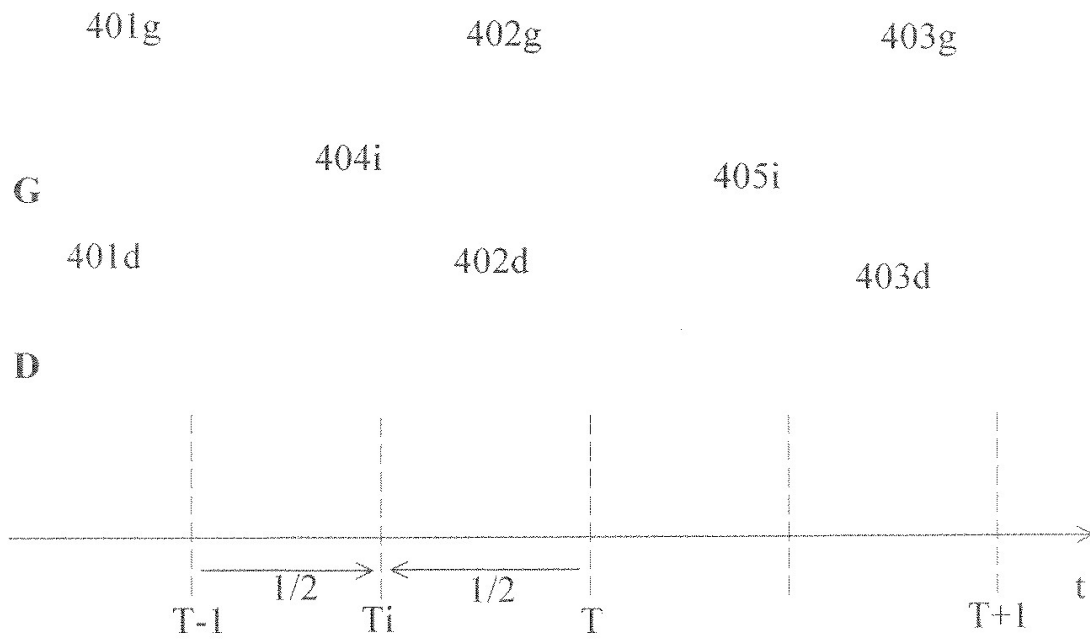


Fig 4A

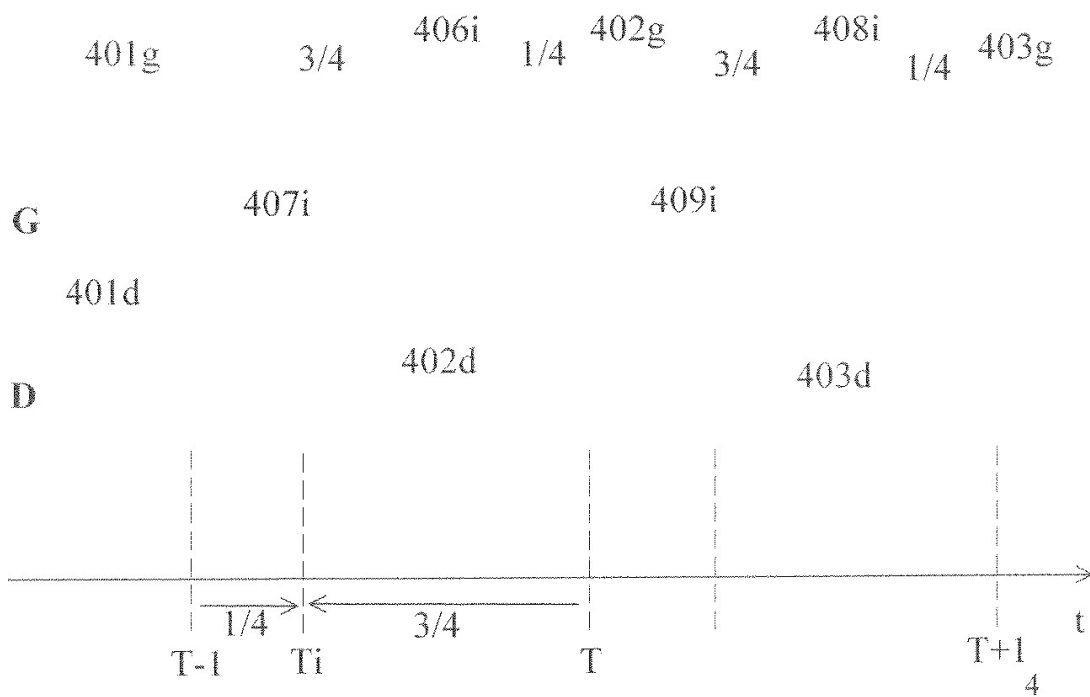


Fig 4B

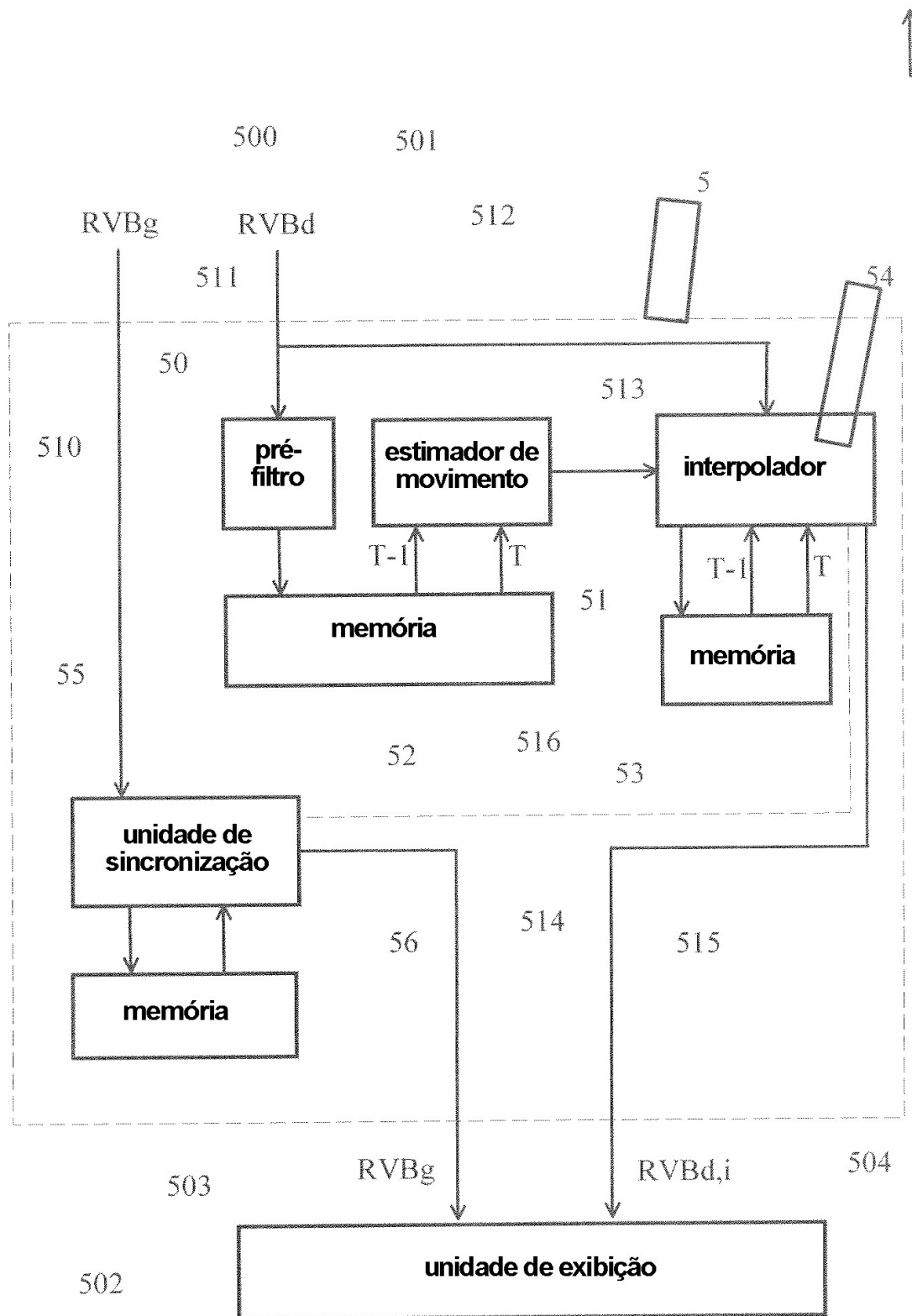


Fig 5



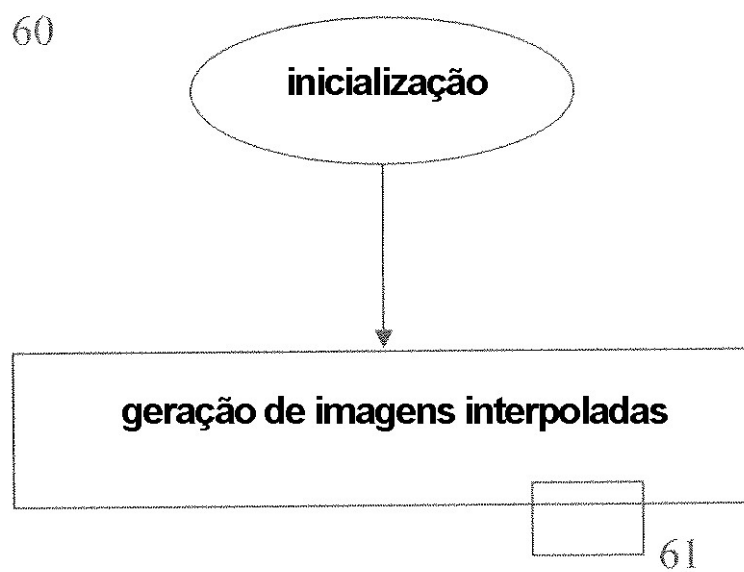


Fig 6

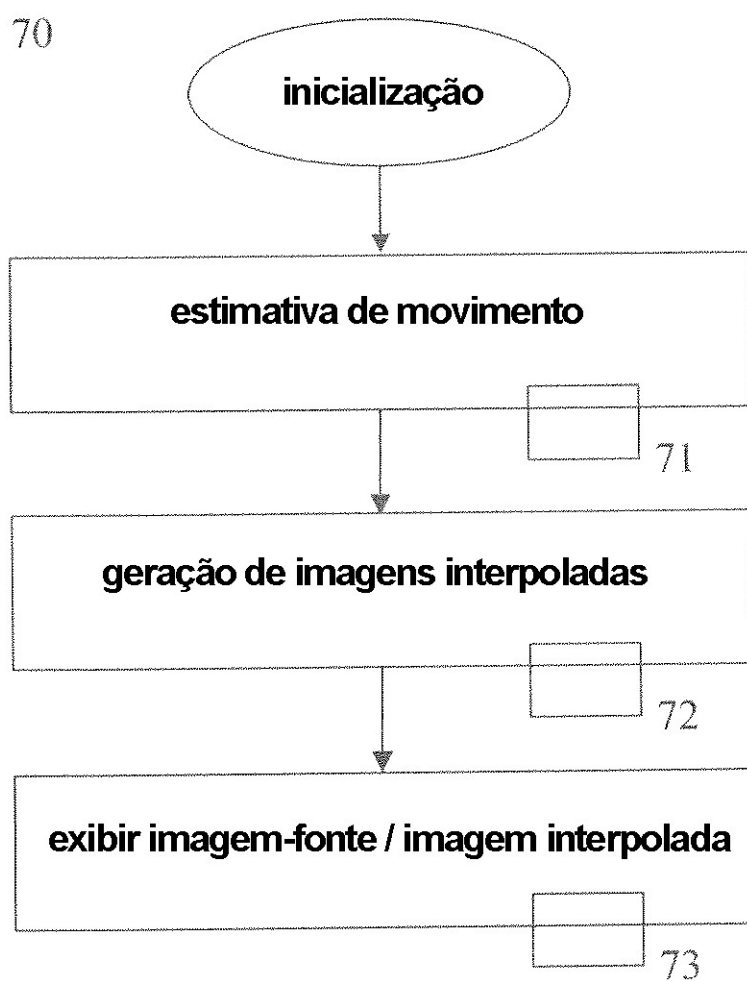


Fig 7