



(22) Data de Depósito: 12/02/2010  
(43) Data da Publicação: 24/01/2012  
(RPI 2142)



**(51) Int.Cl.:**  
H04N 13/02  
G06T 19/20  
G01S 13/00  
H04N 15/00

**(72) Inventor(es):** Clive Henry Gillard, Robert Mark Stefan Porter

Diagrama de uma estação de trabalho de processamento de conteúdo (10) que recebe dados de uma rede de sensores (30). A rede de sensores inclui uma estação base (20) e vários sensores (801, 803, 805) distribuídos em uma área (31a, 31b, 31c, 31d). As setas indicam a transmissão de dados dos sensores para a estação base e para a estação de trabalho de processamento de conteúdo.

“DISPOSITIVO, SISTEMA E MÉTODO DE PROCESSAMENTO DE IMAGEM, E SOFTWARE DE COMPUTADOR”

## **FUNDAMENTOS DA INVENÇÃO**

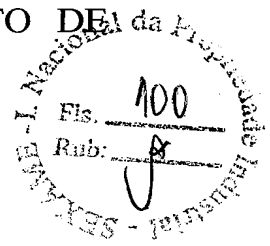
### **Campo da Invenção**

A presente invenção relaciona-se a um dispositivo e método de processamento de imagem.

### **Descrição da Técnica Anterior**

Geração de imagem estereoscópica na qual um par de imagens pode ser visualizado junto, de modo a produzir uma aparência de uma imagem tridimensional tem sido popular por mais de um século. Para criar uma ilusão de que uma imagem é tridimensional (3D), duas imagens ligeiramente diferentes podem ser visualizadas juntas, de tal modo que uma das imagens é visualizada pelo olho esquerdo do usuário e a outra imagem é visualizada pelo olho direito do usuário. Desde que as duas imagens correspondem a duas vistas ligeiramente diferentes da mesma cena (por exemplo, cada imagem no par sendo como se fosse vista a partir do olho esquerdo do usuário e olho direito do usuário respectivamente), o cérebro do usuário engana o usuário para pensar que o par de imagens é uma imagem tridimensional, quando as imagens são visualizadas de uma maneira adequada. Um objeto dentro das imagens aparecerá a uma profundidade aparente do visor, que é dependente de uma quantidade de desvio entre a imagem esquerda correspondente a aquele objeto e a imagem direita correspondente a aquele objeto.

No sentido de tentar e assegurar que cada olho veja a imagem que se pretende que seja visualizada por aquele olho, muitas técnicas são conhecidas. Em algumas técnicas, cada imagem do par estéreo pode ser reproduzida de tal modo a ser separável da outra imagem. Por exemplo, a imagem esquerda deveria ser visualizada próxima à imagem direita e um visualizador adequado tal como um estereoscópio usado para ver as imagens.



Este método de visualização de imagens foi usado nas formas mais antigas de imagens 3D.

Alternativamente, uma imagem esquerda destinada a ser visualizada pelo olho esquerdo pode ser polarizada na direção vertical, enquanto uma imagem direita destinada a ser visualizada pelo olho direito pode ser polarizada na direção horizontal. A imagem esquerda e a imagem direita são então superpostas uma à outra. Usando óculos apropriados nos quais uma polarização de cada lente corresponde à imagem desejada a ser visualizada, a imagem correta para cada olho será visualizada pelo usuário.

Muitos outros sistemas para exibição e visualização de imagens são também conhecidos tais como polarização circular, filtros coloridos (por exemplo, imagens tridimensionais (*anaglyph*) vermelho/ciano), hologramas tridimensionais (*chromadepth*) e óculos 3D (*anachrome*); estes sistemas requerem que um usuário use um par apropriado de óculos. Em outros sistemas, tais como visores que usam um sistema de visor lenticular, não é necessário que o usuário use quaisquer óculos no sentido de visualizar uma imagem tridimensional.

Várias técnicas para gerar um par estéreo de imagens são conhecidas. Um sistema comum para produzir um par de imagens para visualização juntas como um par estéreo de imagens compreende duas câmeras deslocadas uma da outra de modo a capturar duas imagens diferentes da cena. Uma câmera é usada para capturar a imagem da esquerda e a outra câmera é usada para capturar a imagem da direita. As duas imagens capturadas podem então ser visualizadas juntas como uma imagem 3D, usando uma das técnicas descritas acima.

Recentemente, com o advento de câmeras de vídeo de alta definição, e radiodifusão de televisão, alguns radiodifusores têm usado um par de câmeras de vídeo de alta definição para capturar imagens de alta definição de eventos esportivos tais como jogos de futebol e lutas de box para

reprodução como imagens 3D. Entretanto, arranjos de câmeras de vídeo de alta definição atuais para capturar imagens 3D são muito dispendiosos.

Adicionalmente, más combinações de luminância e crominância de imagem entre as duas câmeras pode causar desconforto e dores de cabeça para um usuário, pois o cérebro tentará fazer as duas imagens aparecerem como sendo a mesma.

Técnicas para gerar imagens 3D a partir de imagens bidimensionais (2D) são também conhecidas. Por exemplo, a imagem 2D pode ser considerada ser a imagem esquerda para visualização pelo olho esquerdo. Para gerar a imagem direita, um operador pode selecionar um objeto dentro de uma cena, decidir a qual profundidade aquele objeto deveria aparecer na cena e aplicar uma quantidade de deslocamento de imagem para aquele objeto, dentro daquela cena, de modo a gerar a imagem direita. Um exemplo de tal processo é denominado *Dimensionalization*® que é implementada pela In-Three Inc. Entretanto, no sentido de criar realisticamente uma imagem 3D a partir de uma imagem 2D, um operador pode ter que usar seu julgamento artístico relativo à profundidade de objetos dentro da cena. Adicionalmente, cada quadro de uma seqüência de vídeo ou filme ainda precisa ser marcada individualmente pelo operador de modo a definir os objetos e suas respectivas profundidades dentro da imagem. Além de ser dispendioso, este processo pode consumir muito tempo e de trabalho intenso, significando então que é provável que seja impossível criar vídeo 3D em tempo real a partir de imagens de vídeo capturadas por uma única câmera.

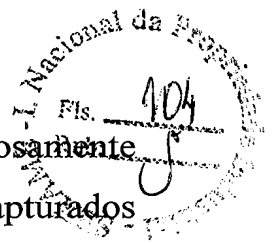
A presente invenção busca aliviar ou mitigar os problemas acima.

### **SUMÁRIO DA INVENÇÃO**

Em um primeiro aspecto, é provido um dispositivo de processamento de imagem compreendendo: meio de recepção operável para receber, de uma câmera, uma imagem capturada correspondente a uma

imagem de uma cena capturada pela câmera, a cena contendo pelo menos um objeto; meio de determinação operável para determinar uma distância entre o objeto dentro da cena e uma posição de referência definida com respeito à câmera; e meio de geração operável para detectar uma posição do objeto dentro da imagem capturada, e para gerar uma imagem modificada a partir da imagem capturada, com base nas características de imagem dentro da imagem capturada que correspondem ao objeto na cena, onde: o meio de geração é operável para gerar a imagem modificada, deslocando a posição do objeto capturado dentro da imagem modificada com respeito à posição determinada do objeto dentro da imagem capturada por uma quantidade de deslocamento que é dependente da distância entre a posição de referência e o objeto na cena, de tal modo que, quando a imagem modificada e a imagem capturada são visualizadas juntas como um par de imagens em um visor, o objeto capturado aparece como posicionado a uma distância predeterminada do visor.

Em um segundo aspecto, é provido um método de processamento de imagem, compreendendo: receber, de uma câmera, uma imagem capturada correspondente a uma imagem de uma cena capturada pela câmera, a cena compreendendo pelo menos um objeto; determinar uma distância entre o objeto dentro da cena e uma posição de referência definida com relação à câmera, detectar uma posição do objeto dentro da imagem capturada; e gerar uma imagem modificada a partir da imagem capturada, com base em características de imagem dentro da imagem capturada que corresponde ao objeto na cena, deslocando a posição do objeto capturado dentro da imagem modificada, com relação à posição determinada do objeto dentro da imagem capturada por uma quantidade de deslocamento de objeto que é dependente da distância entre a posição de referência e o objeto na cena, de tal modo que, quando a imagem modificada e a imagem capturada são visualizadas juntas como um par de imagens em um visor, o objeto capturado aparece como posicionado a uma distância predeterminada do visor.



Realizações da presente invenção permitem vantajosamente cobertura de vídeo, por exemplo, de eventos esportivos a serem capturados por uma câmera de alta definição e então processados de modo a formar imagens 3D. Conseqüentemente, realizações da invenção podem derivar uma quantidade de deslocamento para cada objeto detectado dentro de uma cena, com base em uma distância detectada entre uma posição de referência (tal como a posição da câmera) e o objeto. Quando a imagem capturada (por exemplo, imagem esquerda) e a imagem modificada (por exemplo, imagem direita) são visualizadas juntas como um par de imagens em um visor, o objeto aparecerá como posicionado a uma distância predeterminada do visor, isto permite vantajosamente que imagens 3D sejam geradas sem a necessidade de pares estéreo dispendiosos de câmeras ou operações de pós processamento que consomem tempo.

Ainda mais, deslocando a posição de um objeto dentro de uma imagem modificada com respeito a uma posição detectada do objeto dentro da imagem capturada de uma quantidade de deslocamento que é dependente de uma distância determinada pela estação de trabalho a partir de dados tais como dados de imagem ou dados de distância, imagens 3D podem ser geradas imagens de material não editado de vídeo 2D substancialmente em tempo real. O termo “substancialmente” é destinado a não só capturar em tempo real e quase em tempo real, como também dentro de um período definido de tempo tal como, em realizações, menor que a duração do evento de esportes ou qualquer período dentro daquele evento de esportes, por exemplo, uma metade de um jogo de futebol ou um quarto de um jogo de Futebol Americano.

Vários aspectos adicionais e características da presente invenção são definidos nas reivindicações anexas.

### **BREVE DESCRIÇÃO DOS DESENHOS**

Os acima e outros objetivos e características e vantagens da invenção serão aparentes a partir da seguinte descrição detalhada de

realizações ilustrativas que deve ser lida em conexão com os desenhos que a acompanham, nos quais:

Figura 1 é um diagrama esquemático de um sistema de rastreamento de objeto de acordo com um exemplo da presente invenção;

Figura 2 é um fluxograma de um método de rastreamento de objeto de acordo com exemplos da presente invenção;

Figuras 3A e 3B são diagramas esquemáticos de rastreamento de objeto de acordo com um exemplo da presente invenção;

Figura 4 é um diagrama esquemático de detecção de oclusão de acordo com um exemplo da presente invenção;

Figura 5 é um fluxograma de um método de rastreamento de objeto e detecção de oclusão de acordo com um exemplo da presente invenção;

Figuras 6A e 6B são diagramas esquemáticos de rastreamento de objeto e detecção de oclusão de acordo com um exemplo da presente invenção;

Figuras 7a a 7d mostram diagramas esquemáticos de exemplos de um arranjo de visualização para exibir imagens 3D a um usuário, de tal modo que um objeto pode parecer estar posicionado a uma distância de um visor;

Figura 8 é um diagrama esquemático de um sistema para determinar a distância entre uma câmera e objetos dentro de um campo de visão da câmera, de acordo com realizações da presente invenção;

Figura 9 é um diagrama esquemático de um sistema para determinar a distância entre uma câmera e objetos dentro de um campo de visão da câmera, de acordo com realizações da presente invenção;

Figura 10 mostra um sistema para visualizar imagens de acordo com realizações da presente invenção, de tal modo que as imagens podem ser visualizadas como imagens tridimensionais por um usuário;

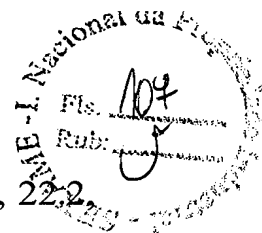
Figura 11 mostra um diagrama esquemático de um grupo de jogadores em campo; e

Figura 12 mostra um fluxograma de um método de processamento de imagem de acordo com realizações da presente invenção.

### **DESCRIÇÃO DAS REALIZAÇÕES PREFERIDAS**

Um dispositivo de processamento de imagem e método de processamento de imagem são descritos. Na seguinte descrição, um número de detalhes específicos são apresentados no sentido de prover um entendimento direto de realizações da presente invenção. Será aparente, entretanto, a uma pessoa especialista na técnica, que estes detalhes específicos não necessitam ser empregados para praticar a presente invenção. Inversamente, detalhes específicos conhecidos da pessoa especialista na técnica são omitidos para fins de clareza na apresentação das realizações.

Figura 1 mostra um diagrama esquemático de um sistema de rastreamento de objeto de acordo com exemplos da presente invenção. No exemplo mostrado na Figura 1, os objetos a serem rastreados são jogadores de futebol (não mostrados) em um campo de futebol 30. Imagens de vídeo de alta definição (HD) (1920 por 1080 pixéis) do campo de futebol 30 são capturadas por uma ou mais câmeras de alta definição. Embora exemplos da presente invenção possam ser usados para rastrear objetos em imagens de vídeo de mais de uma câmera, em alguns exemplos somente uma única câmera é usada. Como será verificado, câmeras HD são dispendiosas, de tal modo que usar somente uma única câmera pode reduzir uma despesa requerida para implementar sistemas que utilizam a presente técnica. Entretanto, usar somente uma única câmera provê somente uma única vista bidimensional de uma cena dentro da qual os objetos são dispostos. Como resultado, rastreamento dos objetos dentro da cena representada pelas imagens de vídeo pode ser mais difícil, porque eventos de oclusão, nos quais um objeto obscurece outro, são mais prováveis. Tal exemplo de câmera única



é mostrado na Figura 1, embora, conforme ilustrado pela câmera 22.1, 22.2, opcionalmente duas câmeras possam ser usadas, cada uma apontando para uma metade diferente do campo de futebol.

Na Figura 1, uma câmera de vídeo 20 é disposta em um ponto fixo dentro do estádio de futebol e arranjada para comunicar sinais representando imagens de vídeo capturadas pela câmera 20 a uma estação de trabalho de processamento de conteúdo 10 que realiza processamento de imagem e outras operações de modo a rastrear a posição dos jogadores no campo, em relação ao tempo. Dados representando a posição dos jogadores em relação ao tempo são então conectados, de tal modo que metadados e estatísticas de jogo podem ser gerados, tais como a extensão de tempo que um jogador particular gasta em uma parte do campo, até onde cada jogador corre, e similares. Os dados representando a posição dos jogadores com relação ao tempo formam dados de trajeto para cada jogador, que se relacionam ao trajeto que cada jogador percorreu dentro das imagens de vídeo. Os dados de trajeto são gerados com respeito a um modelo tridimensional do campo de futebol (plano do objeto) no sentido de prover informação associada ao movimento dos jogadores com respeito a sua posição no campo, que não é prontamente aparente a partir das imagens de vídeo (bidimensionais). Estes dados de trajeto gerados podem então ser usados para reforçar uma experiência de visualização para um espectador quando material não editado do jogo de futebol é transmitido via um meio adequado ao espectador ou para auxiliar um treinador ao comandar o time de futebol. O rastreamento de objetos tais como jogadores no campo de futebol 30 será descrito em mais detalhe abaixo.

Em realizações da presente invenção, a estação de trabalho de processamento de conteúdo 10 usa um processador de célula desenvolvido em conjunto pela Sony®, Toshiba® e IBM®. A natureza paralela do processamento de célula torna-o particularmente adequado para realizar

tarefas de processamento computacionalmente intensas, tais com  
processamento de imagem, reconhecimento de imagem e rastreamento de  
objeto. Entretanto, uma pessoa especialista na técnica verificará que qualquer  
estação de trabalho adequada e unidade de processamento pode ser usada para  
5 implementar realizações da presente invenção.

Será verificado que o sistema e método de rastreamento de  
objeto de acordo com exemplos da presente invenção não precisa ser limitado  
a rastreamento de jogadores em um campo de futebol. Por exemplo, jogadores  
de outros times de esportes tais como rúgbi, críquete, Futebol Americano,  
10 hóquei no gelo, basquete, e similares podem ser rastreados. Adicionalmente,  
objetos tais como bolas, “disco” de hóquei no gelo ou veículos tais como  
carros de corrida e similares podem ser rastreados de acordo com exemplos da  
presente invenção.

De acordo com a presente técnica, as imagens de vídeo que  
15 são geradas usando a câmera de vídeo HD 20 são arranjadas para capturar a  
vista de todo o campo, de tal modo que os jogadores no campo podem ser  
rastreados. Então, todo o campo é capturado a partir de uma posição estática  
da câmera de vídeo HD 20, embora conforme mencionado acima, mais de  
uma câmera pudesse ser usada, no sentido de capturar todo o campo. Em um  
20 exemplo, conforme mencionado acima, as duas câmeras 22.1 e 22.2 podem  
ser usadas cada uma direcionada a diferentes metades do campo. Neste  
exemplo, as imagens de vídeo geradas por cada câmera podem ser  
“costuradas” pela estação de trabalho de processamento de conteúdo 10  
conforme descrito no Pedido de Patente do Reino Unido No. 0624410.7, de  
25 modo a formar imagens de vídeo de alta resolução. Neste exemplo, após  
suportar o processo de “costura”, a saída do agrupamento de câmera pode ser  
considerada como uma única imagem de ultra alta resolução.

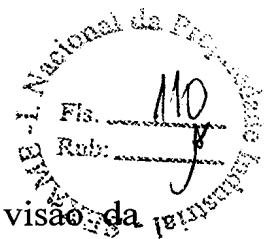
As vantagens do arranjo de ultra alta resolução são numerosas,  
incluindo a capacidade de destacar características particulares de um jogador,

sem ter que aproximar opticamente e portanto afetar a imagem global do estádio. Adicionalmente, o rastreamento automático de um objeto é facilitado porque o segundo plano do evento é estático e há uma resolução de tela mais alta do objeto a ser rastreado.

5 Rastreamento de objeto de acordo com exemplos da presente invenção serão agora descritos com referência às Figuras 2, 3 e 4.

Figura 2 mostra um fluxograma de um método de rastreamento de objeto de acordo com exemplos da presente invenção. No sentido de rastrear um objeto, um modelo de segundo plano é construído a partir das partes do vídeo recebido que são detectadas como sendo substancialmente estáticas ao longo de um número predeterminado de quadros. Em uma primeira etapa S30 a imagem de vídeo recebida da câmera 20, que representa a quadra de futebol, é processada para construir o modelo de segundo plano da imagem. O modelo de segundo plano é construído no sentido de criar uma máscara de primeiro plano que auxilia a identificar e rastrear os jogadores individuais. O modelo de segundo plano é formado na etapa S30, determinando para cada pixel uma média dos pixéis e uma variância dos valores de pixel entre quadros sucessivos, no sentido de construir o modelo de segundo plano. Então, em quadros sucessivos nos quais o valor médio dos pixéis não muda muito, estes pixéis podem ser identificados como pixéis de segundo plano, no sentido de identificar a máscara de primeiro plano.

Tal segmentação segundo plano/primeiro plano é um processo que é conhecido no campo de processamento de imagem e a presente técnica utiliza um algoritmo descrito no documento por Manzanera e Richefeu, intitulado “A robust and Computationally Efficient Motion Detection Algorithm Based on  $\Sigma-\Delta$  Background Estimation”, publicado nos documentos do ICVGIP, 2004. Entretanto, a presente técnica não deveria ser considerada como limitada a esta técnica conhecida e outras técnicas para gerar uma máscara de primeiro plano com respeito a um modelo de segundo



plano para uso no rastreamento, são também conhecidas.

Será verificado que, no caso em que o campo de visão da câmera de vídeo abrange parte da torcida, é improvável que esta seja incluída no modelo de segundo plano, pois estará provavelmente se movendo. Isto é desejável porque é provável aumentar uma carga de processamento no processador de célula ao realizar o rastreamento de objeto, bem como ser desnecessário, pois é improvável que a maioria dos transmissores de esportes esteja interessada em rastrear pessoas na multidão.

Em um exemplo da presente invenção, o modelo de segundo plano é construído no início do jogo e pode até ser feito antes dos jogadores entrarem em campo. Adicionalmente, o modelo de segundo plano pode ser recalculado periodicamente através do jogo, de modo a levar em conta quaisquer mudanças na condição de iluminação, tais como sombras que podem variar através do jogo.

Na etapa S40, o modelo de segundo plano é subtraído da imagem de entrada a partir da câmera para identificar áreas de diferença. Então, o modelo de segundo plano é subtraído da imagem e a imagem resultante é usada para gerar uma máscara para cada jogador. Na etapa S45, um limite é criado com respeito aos valores de pixel em uma versão da imagem que resulta quando o modelo de segundo plano foi subtraído. O modelo de segundo plano é gerado determinando primeiramente a média dos pixels através de uma série de quadros das imagens de vídeo. A partir de valores médios de cada um dos pixels, a variância de cada um dos pixels pode ser calculada a partir dos quadros das imagens de vídeo. A variância dos pixels é então usada para determinar um valor limite, o que variará para cada pixel através de todos os pixels das imagens de vídeo. Para pixels que correspondem a partes da imagem, onde a variância é alta, tais como partes que incluem a torcida, o limite pode ser ajustado para um valor alto, ao passo que as partes da imagem que correspondem à quadra terão um limite mais

baixo, uma vez que a cor e o conteúdo da quadra serão consistentemente os mesmos, independente da presença dos jogadores. Então, o limite determinará se um elemento de primeiro plano está presente e, portanto uma máscara de primeiro plano pode ser identificada correspondentemente. Na etapa S50 uma probabilidade de forma baseada em uma correlação com um modelo de forma humana média, é usada para extrair uma forma dentro da máscara de primeiro plano. Adicionalmente, características de cor são extraídas da imagem, no sentido de criar uma máscara de probabilidade de cor, no sentido de identificar o jogador, por exemplo, a partir da cor da camisa do jogador. Então, a cor da camisa de cada time pode ser usada para diferenciar os jogadores um do outro. Para esta finalidade, a estação de trabalho de processamento de conteúdo 10 gera modelos de cor na dependência das cores conhecidas de cada conjunto de time de futebol. Então, a cor das camisas de cada time é requerida, a cor da camisa do goleiro e a do juiz. Entretanto, será verificado que outros modelos de cor adequados e/ou processos de coincidência de modelo poderiam ser usados.

Retornando à Figura 2, na etapa S50 a estação de trabalho de processamento de conteúdo 10 compara cada um dos pixéis de cada modelo de cor com os pixéis correspondentes à região da camisa da imagem do jogador. A estação de trabalho de processamento de conteúdo então gera um valor de probabilidade que indica uma similaridade entre pixéis do modelo de cor e os pixéis selecionados, para formar uma probabilidade de cor baseada na distância no espaço de cor do valor de saturação de matiz (HSV) a partir dos modelos de cor de time e quadra. Em adição, uma probabilidade de forma é usada para localizar os jogadores, o que é baseado na correlação com um modelo de forma humana média. Adicionalmente, uma probabilidade de movimento é baseada na distância da posição prevista por um estimador de quadrados mínimos recursivo usando posição de partida, velocidade e parâmetros de aceleração.

A criação de máscaras de jogador é ilustrada na Figura 3A. Figura 3A mostra uma visualização de câmera 210 do campo de futebol gerada pela câmera de vídeo 20. Conforme já explicado, a quadra 30 faz parte do modelo de segundo plano, enquanto os jogadores 230, 232, 234, 236, 238, 240 deveriam fazer parte da máscara de primeiro plano conforme descrito acima. Caixas de contorno de jogadores são mostradas como linhas pontilhadas em torno de cada jogador.

Até então, as etapas S30, S40 e S50 são efetuadas com respeito ao processamento de imagem de câmera. Tendo planejado a máscara de primeiro plano, o rastreamento de jogador é efetuado após classificar primeiramente as trilhas de jogador pela proximidade da câmera na etapa S55. Então, os jogadores que são identificados com estando mais próximos à câmera são processados primeiramente, no sentido de eliminar estes jogadores do processo de rastreamento. Na etapa S60, posições de jogadores são atualizadas de modo a minimizar probabilidades de forma, cor e movimento. Na etapa S70 uma máscara de oclusão é construída, a qual exclui regiões de imagem já conhecidas a serem cobertas por outras trilhas de jogador mais próximas. Isto assegura que jogadores parcialmente ou totalmente em oclusão por outros jogadores possam somente ser coincidentes com regiões de imagem visível. A máscara de oclusão melhora a confiabilidade de rastreamento pois reduz a incidência de fusão de trilha (onde duas trilhas seguem o mesmo jogador após um evento de oclusão). Este é um problema particular quando muitos dos alvos parecem os mesmos, porque não podem ser distinguidos (facilmente) pela cor. A máscara de oclusão permite que pixels sejam designados a um jogador próximo e excluídos do jogador adicional, evitando que ambas as trilhas coincidam com o mesmo conjunto de pixels e então mantendo suas identidades separadas.

Segue-se então um processo de rastrear cada jogador, extraindo as características providas dentro da imagem da câmera e mapeando

estas sobre um modelo 3D conforme mostrado na Figuras 3A e 3B. Então, para corresponder a uma posição dentro da imagem 3D produzida pela câmera, uma posição 3D é designada a um jogador que maximiza probabilidades de forma, cor e movimento. Como será explicado brevemente a seleção e mapeamento do jogador a partir da imagem 2D sobre o modelo 3D será modificada se um evento de oclusão tiver sido detectado. Para auxiliar o mapeamento a partir da imagem 2D para o modelo 3D na etapa S65, os jogadores a serem rastreados são inicializados para o efeito de que picos na probabilidade de forma e cor sejam mapeados na seleção de jogadores mais apropriada. Deveria ser enfatizado que a inicialização, que é efetuada na etapa S65 é somente realizada uma vez, tipicamente no início do processo de rastreamento. Para uma boa inicialização do sistema, os jogadores deveriam ser bem separados. Após inicialização, quaisquer erros no rastreamento dos jogadores são correlacionados automaticamente de acordo com a presente técnica, o que não requer intervenção manual.

No sentido de efetuar rastreamento no modelo 3D a partir das posições de imagem 2D, é executada uma transformação pelo uso de uma matriz de projeção P. O rastreamento requer que posições de imagem 2D possam ser relacionadas a posições dentro do modelo 3D. Esta transformação é obtida pelo uso de uma matriz de projeção (P). Um ponto no espaço 2D equivale a uma linha no espaço 3D:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & P_{03} \\ P_{10} & P_{11} & P_{12} & P_{13} \\ P_{20} & P_{21} & P_{22} & P_{23} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \\ w \end{bmatrix}$$

Um ponto em um espaço 2D equivale a uma linha no espaço 3D porque uma terceira dimensão, que é a distancia da câmera, não é conhecida e, portanto, aparecerá correspondendo a uma linha através do modelo 3D. uma altura dos objetos (jogadores) pode ser usada para determinar a distância da câmera. Um ponto no espaço 3D é obtido

selecionando um ponto juntamente com a linha que fica a uma altura fixa acima do nível do solo (a altura humana média). A matriz de projeção  $P$  obtida a priori, uma vez por câmera, antes da coincidência por um processo de calibração de câmera no qual características físicas da quadra, tais como os cantos 31a, 31b, 31c, 31d da quadra 30 são usados para determinar os parâmetros da câmera, o que pode, portanto, auxiliar no mapeamento da posição 2D dos jogadores que tenham sido identificados no modelo 3D. Esta é uma técnica conhecida, usando métodos estabelecidos. Em termos de parâmetros físicos, a matriz de projeção  $P$  incorpora o nível de *zoom* da câmera, centro focal, posição 3D e vetor de rotação 3D (para onde está apontando).

O algoritmo de rastreamento efetuado na etapa S60 é escalável e pode operar em uma ou mais câmeras, requerendo somente que todos os pontos na quadra estejam visíveis de pelo menos uma câmera (com uma resolução suficiente).

Em adição à coincidência de cor e forma, a etapa S60 inclui um processo no qual o movimento do jogador sendo rastreado é também incluído no sentido de identificar corretamente cada um dos jogadores com uma probabilidade maior. Então, o movimento relevante de jogadores entre quadros pode ser determinado em termos de um movimento relevante e em uma direção. Então, o movimento relativo pode ser usado para quadros subseqüentes, para produzir uma região de busca para identificar um jogador particular. Adicionalmente, conforme ilustrado na Figura 3B, o modelo 3D do campo de futebol pode ser aumentado com as linhas 230.1, 232.1, 234.1, 236.1, 238.1, 250.1 que estão posicionadas em relação à indicação gráfica da posição dos jogadores, para refletir a direção relativa de movimento dos jogadores no campo de futebol.

Na etapa S70, uma vez que a posição relativa dos jogadores foi identificada no modelo 3D, então esta posição é correspondentemente

projetada de volta na visualização da imagem 2D da quadra de futebol e um contorno relativo é projetado em torno do jogador, a partir de sua posição no modelo 3D. também na etapa S70, o contorno relativo em torno do jogador é então adicionado à máscara de oclusão para aquele jogador.

5                    Figura 3B mostra uma vista plana de um modelo virtual 220 do campo de futebol. No exemplo mostrado na Figura 3B, os jogadores 230, 232 e 234 (do lado esquerdo do campo) foram identificados pela estação de trabalho de processamento de conteúdo 10 como usando uma camisa de futebol de cor diferente dos jogadores 236, 238 e 240 (do lado direito do campo) indicando então que estão em times diferentes. Diferenciar os jogadores deste modo faz a detecção de cada jogador após um evento de oclusão mais fácil, pois podem ser facilmente distinguidos um do outro pela cor de suas roupas.

15                    Referindo-se de volta à Figura 2, em uma etapa S60, a posição de cada jogador é rastreada usando técnicas conhecidas tais como filtragem de Kalman, embora seja verificado que outras técnicas adequadas podem ser usadas. Então, rastreamento tem lugar em ambas visualização de câmera 210 e modelo virtual 220. Em um exemplo da presente invenção, predição de velocidade realizada pela estação de trabalho de processamento de conteúdo 10 usando a posição dos jogadores no modelo virtual 220 é usada para auxiliar o rastreamento de compreender adicionalmente jogador na visualização de câmera 210.

25                    As etapas S60 e S70 são repetidas até que todos os jogadores tenham sido processados, então o processamento avança para a etapa S60, ao passo que, se o processamento tiver terminado, então termina em S80.

Conforme mostrado na Figura 2, o método ilustrado inclui uma etapa adicional S85, que pode ser requerida se imagens são produzidas por mais de uma câmera. Deste modo, as etapas de processo S30 a S80 podem ser executadas para as imagens de vídeo a partir de cada câmera. Portanto, de

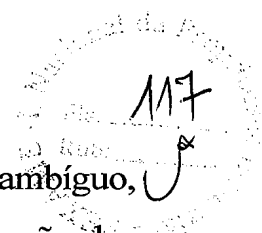
acordo com a etapa S85, cada uma das posições dos jogadores é estimada de acordo com uma probabilidade de detecção a partir de cada câmera. Portanto, de acordo com a etapa S85, cada uma das posições dos jogadores é estimada de acordo com a probabilidade para cada jogador, a partir de cada câmera, e a posição do jogador estimada a partir da mais alta das probabilidades provida por cada câmera, de tal modo que a posição com a probabilidade mais alta para compreender adicionalmente jogador é identificada como a localização para aquele jogador.

Se tiver sido determinado que ocorreu um erro no rastreamento dos jogadores no campo de futebol, então a trilha para aquele jogador pode ser reinicializada na etapa S90. A detecção de um erro no rastreamento é produzida onde uma probabilidade de detecção de um jogador particular é relativamente baixa para uma trilha particular e, conseqüentemente, a trilha é reinicializada.

Um resultado de efetuar o método ilustrado na Figura 2 é gerar dados de trajeto para cada jogador, o que provê uma posição do jogador em cada quadro da imagem de vídeo, que representa um trajeto que aquele jogador segue através do jogo. Então os dados de trajeto provêm posição com respeito ao tempo.

Um problema pode surgir ao rastrear a posição de cada jogador a partir de uma única visualização de câmera, se um jogador obscurece o todo ou uma parte de um outro jogador, como ilustrado na Figura 4.

Figura 4 mostra diversos jogadores 310, 320, 330 e 340 e suas caixas de contorno associadas, conforme indicado pelas linhas tracejadas em torno de cada jogador. Enquanto os jogadores 310 e 340 são claramente distinguíveis um do outro, o jogador 320 obscurece parte do jogador 330. Este é um assim chamado evento de oclusão. Um evento de oclusão pode ocorrer quando todo ou parte de um jogador obscurece todo ou parte de pelo menos um

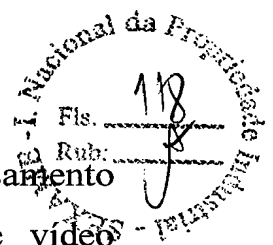


outro jogador com o efeito de que o rastreamento dos jogadores torna-se ambíguo, mesmo depois de outros fatores, tais como o movimento relativo e direção dos jogadores serem levados em conta. Entretanto, será verificado que eventos de oclusão nos quais dois ou mais jogadores estão envolvidos podem ocorrer.

5                    Para detectar um evento de oclusão, a estação de trabalho de processamento de conteúdo 10 detecta se toda ou parte de uma máscara associada a um jogador ocorre na mesma região de imagem como toda ou parte de uma máscara associada a um outro jogador, conforme mostrado na Figura 4. No caso em que jogadores envolvidos em um evento de oclusão  
10                    estão em times opostos, e então possuem camisas de cores diferentes, podem ser facilmente distinguidos e rastreados. Entretanto, após o evento de oclusão, se os jogadores estão ambos do mesmo lado, a estação de trabalho de processamento de conteúdo 10 pode não ser capaz de distinguir qual jogador é quem, particularmente porque seu movimento após um evento de oclusão que  
15                    tenha causado, por exemplo, uma colisão, pode não ser previsível e portanto, pode não rastrear os jogadores corretamente. Como resultado, um trajeto de rastreamento designado a cada jogador pode se tornar trocado.

                    No sentido de resolver uma ambigüidade nos jogadores rastreados, a estação de trabalho de processamento de conteúdo 10 rotula  
20                    todos os jogadores envolvidos no evento de oclusão com as identidades de todos aqueles jogadores envolvidos no evento de oclusão. Então, em um instante mais tarde, se um ou mais dos jogadores se torna facilmente distinguível, a estação de trabalho de processamento de conteúdo 10 usa esta informação para designar novamente as identidades dos jogadores aos  
25                    jogadores corretos, de modo a manter uma gravação de quem é cada jogador. Este processo é descrito em mais detalhe com referência à Figura 5.

                    Figura 5 mostra um fluxograma de um método de rastreamento de objeto e detecção de oclusão de acordo com um exemplo da presente invenção.



Em uma etapa S100, a estação de trabalho de processamento de conteúdo realiza processamento de imagem nas imagens de vídeo capturadas, de modo a extrair uma ou mais características de imagens, conforme descrito acima com referência à Figura 2. As características de imagem extraída são então comparadas com características de imagem que são extraídas de possíveis exemplos dos objetos, de modo a identificar cada objeto. Em um exemplo da presente invenção, jogadores são identificados a partir do número da camisa, como será descrito em mais detalhe abaixo, com referência às Figuras 7 e 8. A estação de trabalho de processamento de conteúdo então gera identificação de objeto para cada objeto, que identifica cada objeto. Alternativamente, em um exemplo da presente invenção, cada objeto (por exemplo, um jogador) é identificado por um operador, via uma interface de operador. A estação de trabalho de processamento de conteúdo então usa os dados de entrada a partir da interface de operador para gerar os dados de identificação de objeto. Entretanto, será verificado pelos especialistas que técnicas de reconhecimento de imagem poderiam ser combinadas com identificação pelo operador, de modo a gerar os dados de identificação do objeto ou que outros métodos poderiam ser usados, tal como reconhecimento de número, que identifica os jogadores pelos números nas costas de suas camisas.

Em uma etapa s105, a estação de trabalho de processamento de conteúdo 10 detecta quaisquer objetos a serem detectados, tais como os jogadores, conforme descrito com referência à Figura 2 acima, na dependência de uma ou mais características de imagem extraídas na etapa s100. Como foi mencionado acima, cada jogador é também rastreado usando ambos modelo virtual 220 e visualização de câmera 210. A estação de trabalho de processamento de conteúdo 10 usa os dados gerados durante o processo de rastreamento para gerar e armazenar dados de trajeto de objeto que descrevem o trajeto que cada objeto segue dentro das imagens de vídeo



recebidas. Os dados de trajeto de objeto tomam a forma de uma amostra das coordenadas x-y do jogador com respeito ao tempo. Em um exemplo da presente invenção, os dados de trajeto tem o formato  $(t_i, x_i, y_i)$ , onde  $t_i$  é o instante de amostra e  $x_i$  e  $y_i$  são as coordenadas x e y do objeto no instante de amostragem  $t_i$ . Entretanto, será verificado que outros formatos de dados de trajeto adequados poderiam ser usados.

Na etapa 115, a estação de trabalho de processamento de conteúdo 10 conecta os dados de identificação de objeto para cada objeto, juntamente com dados de trajeto de objeto que se relacionam ao trajeto que cada objeto seguiu dentro das imagens de vídeo. Os dados conectados são armazenados em um controlador de disco rígido (HDD) ou em memória de acesso randômico dinâmica (DRAM) da estação de trabalho de processamento de conteúdo 10. Isto permite que seja mantida uma gravação de qual jogador foi associado a cada trajeto detectado e rastreado. Os dados conectados podem então ser usados para gerar dados sobre cada jogador e onde estes estiveram durante o jogo. Por exemplo, o tempo que um jogador gastou, em uma área particular do campo poderia ser gerado a partir dos dados armazenados na conexão de associação. Adicionalmente, se por qualquer razão a associação entre o jogador e o trajeto se torna ambígua, por exemplo, como pode acontecer após um evento de oclusão, uma gravação disto pode ser mantida até que a ambigüidade seja resolvida como descrito abaixo. Um exemplo dos dados de identificação de objeto conectados juntamente com os dados de trajeto de objeto é mostrado na Tabela 1 abaixo.

| ObjectID | t        | x        | y        |
|----------|----------|----------|----------|
| A        | $t_1$    | $x_1$    | $y_1$    |
| A        | $t_2$    | $x_2$    | $y_2$    |
| A        | $t_3$    | $x_3$    | $y_3$    |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| A        | $t_i$    | $x_i$    | $y_i$    |

A associação entre os dados de identificação de objeto para

cada objeto e os dados de trajeto de objeto para aquele objeto permitem que cada objeto seja rastreado e identificado de acordo. Nos exemplos descritos acima, cada jogador pode ser rastreado, permitindo portanto que um radiodifusor saiba de qual jogador se trata, mesmo que aquele jogador possa estar distante demais para ser visualmente identificado por um operador ou por reconhecimento de imagem realizado pela estação de trabalho de processamento de conteúdo 10. Isto permite que um radiodifusor incorpore características adicionais e informação baseada nesta associação que um espectador do conteúdo de radiodifusão pode considerar desejável. Em uma etapa s120, a estação de trabalho de processamento de conteúdo 10 detecta se um evento de oclusão ocorreu conforme descrito acima com referência à Figura 4. Se nenhum evento de oclusão é detectado, então o processo retorna à etapa s105, na qual os objetos são detectados. Deste modo, compreender adicionalmente objeto pode ser individualmente rastreado e o trajeto de cada objeto univocamente associado à identidade daquele objeto.

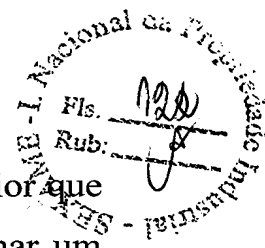
Entretanto, se é detectado um evento de oclusão, então em uma etapa s125, o processador Cell associa os dados de identificação de objeto para cada objeto envolvido no evento de oclusão aos dados de trajeto de objeto para cada objeto envolvido no evento de oclusão. Por exemplo, se dois objetos rotulados A e B são associados a trajetos P e Q, respectivamente, após a detecção de um evento de oclusão envolvendo objetos A e B, o trajeto P será associado a ambos A e B e o trajeto Q será associado a ambos A e B. as associações geradas pela estação de trabalho de processamento de conteúdo 10 após o evento de oclusão são então conectadas conforme descrito acima. Isto permite que os objetos (por exemplo, jogadores) envolvidos no evento de oclusão sejam rastreados sem ter que identificar novamente cada objeto, mesmo se há alguma incerteza de qual jogador é quem. Portanto, uma carga de processamento na estação de trabalho de processamento de conteúdo 10 é reduzida, pois somente aqueles objetos envolvidos no evento de oclusão são

identificados ambigualmente, enquanto objetos não envolvidos máscara de oclusão evento de oclusão podem ainda ser identificados.

Em uma etapa s130, a estação de trabalho de processamento de conteúdo 10 verifica se uma identificação de um ou mais objetos envolvidos no evento de oclusão foi feita, de tal modo que a identidade dos objetos associados aos trajetos gerados pode ser resolvida. A identificação de pelo menos um dos objetos é realizada pela estação de trabalho de processamento de conteúdo, comparando uma ou mais características de imagem associadas a aquele objeto com as características de imagem extraídas dos possíveis exemplos dos objetos. Se nenhuma identificação foi feita, então o processo passa à etapa s105 com os dados de trajeto gerados para cada objeto sendo associados a todos aqueles objetos envolvidos no evento de oclusão.

Entretanto, se uma identificação de um ou mais dos objetos envolvidos no evento de oclusão é detectada como tendo ocorrido, então em uma etapa s135, os dados de trajeto conectados são atualizados para refletir a identificação do objeto que foi positivamente identificado. No exemplo dado acima, o *log* de associação seria atualizado de tal modo que A é associado ao trajeto P e B é associado ao trajeto Q.

Alternativamente, uma identificação de um objeto pode ser realizada por um operador via uma interface de operador, pela estação de trabalho de processamento de conteúdo 10 usando técnicas de reconhecimento de imagem de acordo com exemplos da presente invenção (conforme descrito abaixo) ou por uma combinação das duas técnicas. Entretanto, será verificado que qualquer outra técnica de identificação adequada para distinguir ou identificar cada objeto, poderia ser usada. No caso de reconhecimento de imagem, a estação de trabalho de processamento de conteúdo 10 pode gerar um nível de confiança que indica o quanto a identificação feita pelo processo de reconhecimento de imagem tem probabilidade de estar correta. Em um exemplo da presente invenção, uma



identificação é determinada para estar onde o nível de confiança é maior que um limite predeterminado. Adicionalmente, um operador pode designar um nível de confiança a sua identificação e, se aquele nível de confiança excede um limite predeterminado, então uma identificação é detectada.

5                   Em exemplos da presente invenção, uma história de eventos é gerada, indicando quando os dados de trajeto conectados foram atualizados e isto pode também ser armazenado de modo a atuar como reserva no caso da identificação positiva se revelar incorreta. Por exemplo, uma identificação poderia se mostrar incorreta onde um operador estava convencido de que um

10 jogador que estava distante da câmera de vídeo 20 tinha uma identidade particular, porém o jogador veio mais próximo da câmera de vídeo (permitindo ao usuário ver uma imagem de resolução mais alta do jogador), o operador verifica que houve erro. Neste caso, pode ser usada a interface de operador para superpor sua identificação prévia do jogador, de modo que a

15 estação de trabalho de processamento de conteúdo 10 possa atualizar os dados de trajeto conectados de acordo. No exemplo dado acima, um histórico de evento de identificação pode ser armazenado em um controlador de disco rígido (HDD) ou em memória de acesso randômico dinâmica (DRAM) da

20 estação de trabalho de processamento de conteúdo 10, com dados mostrando que, antes da identificação positiva, o trajeto P usado para estar associado a ambos A e B e o trajeto Q usado para estar associado a ambos A e B.

O histórico de evento de identificação pode também incluir o nível de confiança que foi gerado durante o processo de identificação. Se uma identificação subsequente é feita de um objeto que tem um nível de confiança

25 mais alto que o de uma identificação positiva prévia, então o nível de confiança da identificação subsequente pode ser usado para verificar ou anular a identificação prévia.

Será verificado que, após a detecção de um evento de oclusão, um objeto pode ser identificado a qualquer instante após o evento de oclusão,

de modo a excluir a ambigüidade dos objetos envolvidos no evento de oclusão. Portanto, após a detecção de um evento de oclusão, a estação de trabalho de processamento de conteúdo 10 pode monitorar se uma identificação positiva de um objeto ocorreu como um processo de segundo plano que é executado simultaneamente com as etapas s105 a s125.

Alguns exemplos de rastreamento de objeto e detecção de oclusão de acordo com exemplos da presente invenção serão agora descritos com referência às Figuras 6a e 6b.

No exemplo mostrado na Figura 6a, dois objetos identificados como A e B estão envolvidos em um evento de oclusão 410. Após o evento de oclusão, ambos trajetos de objeto detectados conforme indicado pelas setas, são associados a ambos A e B (AB). Algum tempo depois, o objeto B é positivamente identificado como indicado por AB no trajeto inferior. Esta identificação é então usada para atualizar a associação entre o objeto e os trajetos, de tal modo que o objeto A é associado ao trajeto superior após o evento de oclusão 410 e o objeto B é associado ao trajetos inferior após o evento de oclusão 410.

No exemplo mostrado na Figura 6b, objetos A e B são inicialmente envolvidos em um evento de oclusão 420. Entretanto, antes dos objetos A e B poderem ser positivamente identificados, o objeto associado a ambos A e B no trajeto inferior após o evento de oclusão 420 ser envolvido em um outro evento de oclusão 430 com o objeto C. Conseqüentemente, antes do evento de oclusão 430, não está claro se o objeto no trajeto inferior após o evento de oclusão 420 é o objeto A ou objeto B. Portanto, após o evento de oclusão 430, ambos trajetos superior e inferior que os dois objetos seguem estão associados aos objetos A, B e C (ABC).

Em um instante posterior, o objeto no trajeto inferior após o evento de oclusão 430 é positivamente identificado como sendo o objeto B (ABC). Portanto, o *log* de associação pode ser atualizado de tal modo que o

trajeto superior após o evento de oclusão 430 é associado ao objeto A. Adicionalmente, esta informação pode ser usada para atualizar o *log* de associação, de tal modo que dois objetos envolvidos no evento de oclusão 420 podem ser tornados não ambíguos pois deve ter sido o objeto B que foi envolvido no evento de oclusão 430 como objeto B que foi positivamente identificado como estando associado ao trajeto inferior após o evento de oclusão 430. Conseqüentemente, o *log* de associação pode ser adaptado de tal modo que o trajeto superior após o evento de oclusão 420 é associado ao objeto A e o trajeto inferior após o evento de oclusão 420 associado ao objeto B.

Portanto, exemplos da presente invenção permitem que objetos sejam associados a trajetos rastreados de objetos, embora vários eventos de oclusão possam ter ocorrido antes que um objeto seja positivamente identificado. Adicionalmente. Exemplos da presente invenção permitem que as identidades dos diferentes objetos tenham referência cruzada umas com as outras, de tal modo a permitir que cada trajeto seja associado ao objeto correto.

Em alguns exemplos, dados representando a posição de partida de objetos podem ser usados para inicializar e verificar o rastreamento de objeto. Considerando futebol como um exemplo, é provável que jogadores iniciem um jogo em posições aproximadamente estacionárias no campo de jogo. Cada jogador provavelmente é posicionado dentro de uma distância limite de uma coordenada particular no campo de jogo. As posições de partida podem depender da formação do time, tais como 4-4-2 (quatro na defesa, quatro no meio de campo e dois no ataque) ou 5-3-2, e também de qual time está atacando e qual time está defendendo. Posições similares são prováveis de serem adotadas por jogadores a partir de um chute a gol a partir do solo. Tal posição pode ser usada para iniciar rastreamento de jogador, por exemplo, comparando dados de posição com um esquema do time e informação de

formação. Tal informação de posição pode também ser usada para corrigir a  
informação de trajeto quando um evento de oclusão tiver ocorrido. Usar a  
informação de formação de time é vantajoso porque esta pode ser reiniciada  
por uma operação no decurso de um jogo, se as mudanças na formação do time  
se tornam aparentes, por exemplo, após uma substituição ou expulsão. Isto  
melhorará a precisão e confiabilidade do rastreamento de objeto.

Realizações da presente invenção nas quais uma imagem  
capturada pela câmera 20 é usada para gerar uma imagem modificada, de tal  
modo que a imagem capturada e a imagem modificada podem ser  
visualizadas juntas, de modo a parecer a um usuário como uma imagem  
tridimensional, serão agora descritas com referência às Figuras 7a a 7d e  
Figuras 8 a 12.

Figuras 7a a 7d mostram exemplos de um arranjo de  
visualização para exibir imagens a um usuário, de tal modo que um objeto  
pode aparecer posicionado a uma distância de um visor. Usando óculos  
apropriados, tais como aqueles tendo lentes diferentemente coloridas ou  
polarizadas para cada olho, um usuário pode ver imagens exibidas de tão  
modo que as imagens para visualização por seu olho esquerdo e imagens  
destinadas à visualização por seu olho direito sejam vistas corretamente, de  
modo a prover uma experiência de visualização tridimensional para o usuário.

No sentido de criar uma ilusão de que um objeto é  
tridimensional, uma imagem destinada ao olho esquerdo pode ser desviada ou  
deslocada com respeito a uma imagem destinada ao olho direito, de uma  
quantidade de deslocamento que depende da distância entre o usuário e o  
visor, da distância inter pupilar do usuário (IPD), que é a distância entre as  
pupilas do usuário quando o usuário está focalizando seus olhos no infinito, e  
uma distância desejada da tela na qual um objeto é destinado a ser  
posicionado. A distância aparente entre o usuário e um objeto fornecido é  
referida aqui como profundidade aparente do objeto.

O modo pelo qual a quantidade de deslocamento pode ser calculada é descrita em detalhe abaixo, com referência às Figuras 7a a 7d. Em realizações da invenção, a quantidade de deslocamento é aplicada com respeito à imagem destinada ao olho esquerdo (a imagem da esquerda), embora seja verificado que a quantidade de deslocamento poderia ser aplicada com respeito à imagem da direita destinada ao olho direito.

Figura 7a mostra esquematicamente um visor 705 juntamente com ilustrações de um olho esquerdo do usuário 710L e olho direito 710R. A distância entre os olhos do usuário e o visor 705 é rotulada como  $d_s$  e a distância inter pupilar é rotulada como  $p$  nas Figuras 7a a 7d. Figura 7a também ilustra posições no visor, nas quais características de imagem correspondentes a um objeto devem ser exibidas de modo a prover um efeito 3D. Em particular, uma imagem esquerda 715L destinada a visualização pelo olho esquerdo do usuário 710L e uma imagem direita 715R destinada a visualização pelo olho direito do usuário 710R são ilustradas com respeito ao visor 705. A quantidade de deslocamento entre a imagem esquerda 715L e a imagem direita 715R é indicada como  $i$  nas Figuras 7a a 7d.

Na Figura 7a, as posições da imagem esquerda 715L e da imagem direita 715R são tais que a quantidade de deslocamento é a mesma da distância inter pupilar  $p$  (IPD) e portanto o objeto aparecerá como se estivesse a uma distância infinita do visor 705, conforme indicado pelas linhas paralelas tracejadas na Figura 7a. Entretanto, no sentido de fazer o objeto aparecer a uma distância finita do visor 705, a quantidade de deslocamento deveria ser menor que a distância entre as pupilas do usuário (distância inter pupilar  $p$  (IPD)). Isto é ilustrado na Figura 7b. Tipicamente, a distância inter pupilar é de 52 mm a 70 mm, embora possa variar dependendo do sexo, etnia, idade e similar.

Figura 7b mostra um exemplo onde a quantidade de deslocamento  $i$  é menor que a distância  $p$  entre os olhos do usuário, e portanto

um objeto 720 aparecerá como se o objeto 720 estivesse a uma distância  $d_o$  dos olhos do usuário. A quantidade de deslocamento  $i$  pode ser calculada a partir da equação 1 mostrada abaixo.

$$\left( \frac{d_o - d_s}{d_o} \right) \times p = i \quad \text{Equação 1}$$

Na equação 1,  $d_o$  é a profundidade aparente do objeto,  $d_s$  é a distância entre os olhos do usuário 710L e 710R e o visor 705,  $p$  é a distância entre as pupilas do usuário (distância inter pupilar) e  $i$  é a quantidade de deslocamento.

Se o objeto 720 deve aparecer conforme visualizado no visor 705 na posição do visor (em outras palavras  $d_o = d_s$ ), então a quantidade de deslocamento  $i$  será zero. Esta situação é ilustrada na Figura 7c.

Se o objeto deve aparecer como se estivesse posicionado entre o visor 705 e o usuário (em outras palavras  $d_o < d_s$ ), então a quantidade de deslocamento com respeito à imagem direita 715R deveria ser negativa, conforme ilustrado na Figura 7d. em outras palavras, a imagem direita 715R apareceria ao usuário com estando à esquerda da imagem esquerda 715L, fazendo então com que os olhos do usuário convirjam e focalizem em uma posição entre o usuário e o visor 705.

Em realizações, para reduzir o desconforto e fadiga do olho para o usuário, a quantidade de deslocamento deveria ser a mesma, ou menor que a distância inter pupilar ( $i \leq P$ ). Adicionalmente, para reduzir a fadiga e desconforto do olho para o usuário, a relação da distância aparente do objeto  $d_o$  com respeito à distância do visor  $d_s$  deveria ser maior que uma quantidade predeterminada, tipicamente 0,2. Em outras palavras, em algumas realizações,  $d_o/d_s \geq 0,2$ , de modo a tentar reduzir o desconforto para o olho do usuário. Entretanto, será verificado que outros valores adequados poderiam ser usados.

Realizações da presente invenção nas quais uma distância entre uma câmera e um objeto dentro de uma imagem capturada pela câmera é usada para determinar a quantidade de deslocamento será agora descrita

com referência às Figuras 8 a 12.

Figura 8 é um diagrama esquemático de um sistema para determinar a distância entre uma posição da câmera e objetos dentro de um campo de visão da câmera, de acordo com realizações da presente invenção.

5 O sistema mostrado na Figura 8 é similar ao descrito acima com respeito à Figura 1. Em particular Figura 8 mostra a estação de trabalho de processamento de conteúdo arranjada para se comunicar com a câmera 20, que captura imagens do campo 30. Conforme descrito acima, a estação de trabalho de processamento de conteúdo 10 é operável para analisar as  
10 imagens capturadas pela câmera 20, de modo a rastrear jogadores no campo 30, e determinar sua posição no campo 30. Em algumas realizações, o sistema compreende um detector de distância 810 operável para detectar uma distância entre a câmera 20 e objetos dentro do campo de visão da câmera. O detector de distância 810 e sua operação será descritos em mais detalhe  
15 abaixo.

Em algumas realizações, a estação de trabalho de processamento de conteúdo pode usar os dados de rastreamento e dados de posição para determinar uma distância entre uma posição da câmera 20 e jogadores no campo. Por exemplo, a estação de trabalho de processamento de  
20 conteúdo pode analisar a imagem capturada de modo a determinar uma distância 801a entre uma posição da câmera 20 e um jogador 801, uma distância 803a entre a posição da câmera 20 e um jogador 803, e uma distância 805a entre a posição da câmera 20 e um jogador 805.

Em outras palavras, realizações da invenção determinam a  
25 distância entre o objeto dentro da cena e uma posição de referência definida com respeito à câmera. Nas realizações descritas com referência à Figura 8, a posição de referência está localizada na posição da câmera.

Adicionalmente, em algumas realizações, a estação de trabalho 10 é operável para detectar características de imagens predeterminadas dentro

da imagem capturada, que correspondem a pontos de característica conhecidos dentro da cena. Por exemplo, a estação de trabalho pode analisar a imagem capturada, usando técnicas conhecidas de modo a detectar características de imagem que correspondem a características do campo de futebol, tais como cantos, círculo central, grande área e similares, com base nas posições detectadas dos pontos de característica conhecidos detectados (características de imagem), a estação de trabalho 10 pode então mapear o modelo tridimensional do campo 10 para a imagem capturada, usando técnicas conhecidas. Conseqüentemente, a estação de trabalho 10 pode então analisar a imagem capturada para detectar a distância entre a câmera e o jogador, na dependência da posição detectada do jogador com respeito ao modelo 3D que foi mapeado para a imagem capturada.

Em algumas realizações da invenção, a estação de trabalho de processamento de conteúdo 10 pode analisar as imagens capturadas, de modo a determinar uma posição na qual os pés do jogador estão em contato com o campo. Em outras palavras, a estação de trabalho de processamento de conteúdo 10 pode determinar um ponto de intersecção no qual um objeto, tal como um jogador, coincide com uma superfície planar tal como o campo 30.

Onde um objeto é detectado como coincidente com a superfície planar em mais de um ponto de intersecção (por exemplo, ambos os pés do jogador estão em contato com o campo 30), então a estação de trabalho de processamento de conteúdo 10 é operável para detectar qual ponto de intersecção está mais próximo da câmera 20 e usar aquela distância para gerar a quantidade de deslocamento. Alternativamente, uma distância média de todos os pontos de intersecção detectados para aquele objeto pode ser calculada e usada ao gerar a quantidade de deslocamento. Entretanto, será verificado que outros pontos de intersecção adequados poderiam ser selecionados, tal como um ponto de intersecção mais distante da câmera 20.

Entretanto, em algumas situações, o método para determinar

distância entre a posição da câmera 20 e o objeto dentro da cena, conforme descrito acima, pode causar distorções na aparência da imagem tridimensional. Tais distorções podem ser particularmente aparentes, se a imagem é capturada por uma câmera em ângulo muito largo ou formado esticando imagens capturadas por duas câmeras de alta definição tais como a câmera 22.1 e a câmera 22.2.

Por exemplo, distorções de imagem na imagem tridimensional podem ocorrer se o campo 30 deve ser exibido como imagens tridimensionais nas quais os jogadores e a bola são superpostos. Neste caso, os cantos 31b e 31c aparecerão mais distantes de um ponto central 814 sobre a linha lateral mais próxima à câmera 20. A linha lateral pode então parecer curva, embora a linha lateral seja reta na imagem capturada.

Este efeito pode ser particularmente aparente quando a imagem tridimensional é visualizada em uma tela comparativamente grande tal como uma tela de cinema, este efeito é menos óbvio porque os cantos 31b e 31c são mais prováveis de estarem na visão periférica do espectador. O meio pelo qual o campo pode ser exibido como uma imagem tridimensional será descrito em mais detalhe abaixo.

Um modo possível para equacionar este problema seria gerar uma quantidade de deslocamento apropriada para cada parte da imagem, de modo a compensar quanto à distorção. Entretanto, isto pode ser computacionalmente intensivo, bem como ser dependente de vários parâmetros físicos tais como grau de distorção devido a imagem de ângulo amplo, tamanho de visualização e similar.

Portanto, para reduzir a distorção na imagem tridimensional e tentar assegurar que a parte frontal do campo (isto é, a linha lateral mais próxima da câmera) apareça a uma profundidade constante a partir do visor, especialmente quando a imagem tridimensional deve ser visualizada em um visor relativamente pequeno tal como um monitor de computador ou tela de

televisão, realizações da invenção determinam a distância entre o objeto e uma posição de referência que permanece em uma linha de conferência. A linha de referência é ortogonal ao eixo óptico da câmera, e passa através de uma posição da câmera, e a posição de referência está localizada sobre a linha de referência em um ponto em que uma linha de localização de objeto e a linha de referência se interceptam. A linha de localização de objeto é ortogonal à linha de referência e passa através do objeto. Isto será descrito abaixo com referência à Figura 9.

Figura 9 é um diagrama esquemático de um sistema para determinar a distância entre uma câmera e objetos dentro de um campo de visão da câmera, de acordo com realizações da presente invenção. A realização mostrada na Figura 9 é substancialmente a mesma que a descrita acima com referência à Figura 8. Entretanto, nas realizações mostradas na Figura 9, a estação de trabalho é operável para determinar uma distância entre um objeto e uma linha de referência indicada pela linha tracejada 907.

Conforme mostrado na Figura 9, a linha de referência 907 é ortogonal ao eixo óptico da câmera (isto é, a ângulos retos com o eixo óptico) e passa através da posição da câmera. Adicionalmente, Figura 9 mostra posições de referência 901a, 903a e 905a que permanecem na linha de referência 907.

Por exemplo, a estação de trabalho é operável para determinar uma distância 901 entre a posição de referência 901a e o jogador 801. A posição de referência 901a está localizada na linha de referência 907 onde uma linha de referência de objeto (indicada pela linha pontilhada 901b) para o jogador 801 intercepta a linha de referência 907. Similarmente, a posição de referência 903a está localizada na linha de referência 907 onde uma linha de referência de objeto (indicada pela linha pontilhada 903b) para o jogador 803 intercepta a linha de referência 907. Similarmente, a posição de referência 905a está localizada na linha de referência 907 onde uma linha de referência

de objeto (indicada pela linha pontilhada 903b) para o jogador 803 intercepta a linha de referência 907, e a posição de referência 905a está localizada na linha de referência 907 onde uma linha de referência de objeto (indicada pela linha pontilhada 905b) intercepta a linha de referência 907. As linhas de referência de objeto 901b, 903b e 905b são ortogonais à linha de referência 907 e passam através dos jogadores 801, 803 e 805, respectivamente.

Em algumas realizações, a linha de referência 907 é paralela à linha lateral que une os cantos 31b e 31c de tal modo que, quando uma imagem capturada do campo e uma imagem modificada do campo são visualizadas juntas em um visor de uma maneira adequada, todos os pontos sobre a linha lateral unindo cantos 31b e 31c aparecem como se estivessem a uma distância constante (profundidade) do visor. Isto melhora a aparência da imagem tridimensional, sem ter que gerar uma quantidade de deslocamento que compensa qualquer distorção que possa surgir quando a imagem é capturada usando uma câmera de ângulo amplo ou a partir de uma imagem composta formada combinando imagens capturadas por duas ou mais câmeras. Entretanto, será verificado que a linha de referência não necessita ser paralela à linha lateral, e poderia ser paralela a qualquer outra característica apropriada dentro da cena, ou arranjada com respeito a qualquer outra característica apropriada dentro da cena.

No sentido das imagens serem geradas de tal modo que, quando visualizadas, apareçam como tridimensionais, a estação de trabalho de processamento de conteúdo 10 é operável para detectar uma posição de um objeto tal como um jogador dentro da imagem capturada. O modo pelo qual objetos são detectados dentro da imagem, pela estação de trabalho de processamento de conteúdo 10, é descrito acima com referência à Figura 2. A estação de trabalho de processamento de conteúdo 10 pode então gerar uma imagem modificada a partir da imagem capturada, deslocando a posição do objeto dentro da imagem capturada, da quantidade de deslocamento, de tal

modo que, quando a imagem modificada e a imagem capturada são visualizadas juntas como um par de imagens no visor 705, o objeto aparece como posicionado a uma distância predeterminada do visor, o modo pelo qual a imagem modificada e a imagem capturada podem ser visualizadas juntas, é ilustrado na Figura 10.

Figura 10 mostra um sistema para visualizar imagens de acordo com realizações da presente invenção, de tal modo que as imagens podem ser visualizadas como imagens tridimensionais por um usuário.

Em particular, Figura 10 mostra imagens do jogador 801 e do jogador 803 no visor 705. A imagem capturada pela câmera 20 é usada para visualizar uma imagem esquerda 801L correspondendo ao jogador 801, bem como uma imagem esquerda 803L do jogador 803. As imagens esquerdas são destinadas a serem visualizadas pelo olho esquerdo do usuário, por exemplo, pelo usuário usando um par adequado de óculos polarizados. A estação de trabalho de processamento de conteúdo 10 é operável para gerar uma imagem modificada a partir da imagem capturada, de modo a gerar respectiva imagem direita compreendendo cada objeto, Figura 10 mostra uma imagem direita 801R (indicada pela linha tracejada) correspondendo ao jogador 801, e uma imagem direita 803R (indicada pela linha tracejada) correspondendo ao jogador 803. Por exemplo, quando a imagem esquerda 801L é visualizada juntamente com a imagem direita 801R no visor 705, o jogador 801 aparecerá como se estivesse posicionado a uma distância predeterminada do visor 705.

No sentido de gerar a imagem modificada a partir da imagem esquerda (por exemplo, gerar a imagem direita a partir da imagem esquerda) a estação de trabalho de processamento de conteúdo 10 é operável para gerar uma máscara que corresponde a uma linha externa do objeto, tal como o jogador. Isto é descrito acima com referência à etapa S40 da Figura 2. A estação de trabalho de processamento de conteúdo 10 é então operável para aplicar a quantidade de deslocamento de imagem deslocada a pixels dentro da

máscara, de modo a gerar a imagem modificada (imagem direita). Isto pode ser realizado com respeito a cada objeto que é detectado dentro da imagem capturada.

A quantidade de deslocamento para cada jogador é dependente da distância entre a câmara e o jogador. Por exemplo, conforme mostrado na Figura 8, o jogador 801 está mais próximo da câmara que o jogador 803. Portanto, de acordo com a equação 1 acima, para uma dada distância ( $d_s$ ) entre a tela e o usuário, a quantidade de deslocamento entre a imagem esquerda 801L e a imagem direita 801R correspondendo ao jogador 801 será menor que a quantidade de deslocamento entre a imagem esquerda 803L e a imagem direita 803R correspondendo ao jogador 803. A distância aparente de cada objeto pode ser escalada apropriadamente conforme desejado, por exemplo, de modo a ser exibida em um tamanho particular de visualização.

Será verificado que, em algumas circunstâncias, por exemplo, com jogadores de futebol em um campo de futebol, pode ser indesejável fazer com que um jogador apareça em três dimensões a uma distância do visor 705 que corresponde à distância real da câmara 20, pois isto pode causar uma experiência de visualização desagradável para um usuário. Adicionalmente, pode ser perdido algum efeito tridimensional, se um objeto é fornecido de modo a aparecer a dezenas de metros do visor. Portanto, em realizações da invenção, a estação de trabalho 10 é operável para detectar qual percentagem da imagem capturada na direção vertical é ocupada pelo campo de futebol e escalar a profundidade aparente do objeto de acordo.

Por exemplo, a estação de trabalho de processamento de conteúdo 10 pode detectar uma posição de uma linha lateral do campo de futebol 30 que esteja mais próxima da câmara 20, bem como detectar uma posição de uma linha lateral do campo de futebol 30 que esteja mais distante da câmara 20, com base no mapeamento do modelo 3D para a imagem capturada. A estação de trabalho 10 então gera a quantidade de deslocamento



de acordo, de tal modo que objetos que estão à mesma distância da câmera que a linha lateral mais próxima, aparecem como se estivessem à mesma distância do usuário que o visor.

A distância na qual a linha lateral mais distante aparece do visor pode então ser ajustada pela estação de trabalho de processamento de conteúdo 10 para ser uma distância correspondente a uma altura vertical do visor 705. Entretanto, será verificado que qualquer outro método adequado de escalamento da profundidade aparente do objeto pode ser usado.

Adicionalmente, será verificado que é a distância física entre a imagem direita e a imagem esquerda no visor que faz com que o objeto apareça como se estivesse a uma distância predeterminada do visor. Portanto, em realizações da invenção, a quantidade de deslocamento é inicialmente calculada em unidades físicas de medição, tais como milímetros. Ao gerar a imagem modificada para fornecimento como pixéis no visor 705, o valor da quantidade de deslocamento em milímetros é escalado pela estação de trabalho de processamento de conteúdo 10, na dependência de algum ou todos dentre: o tamanho do visor 705; a resolução do visor em pixéis; e o passo de pixel. Estes parâmetros podem ser armazenados em uma tabela de pesquisa que armazena os parâmetros relevantes para diferentes tipos de visor (por exemplo, por fabricante e número de modelo) ou estes podem ser inseridos por um usuário.

Em algumas realizações, a estação de trabalho de processamento de conteúdo 10 pode fazer com que o visor 705 exiba uma sequência de calibração de imagens que permite que um usuário proveja realimentação através de um meio de entrada adequado de modo que, por exemplo, um objeto aparece no infinito, na distância da tela e distâncias entre infinito e o usuário. Entretanto, será verificado que outros métodos adequados de escalamento das imagens direita e esquerda para saída em um visor podem ser usados.

Conforme descrito acima, em algumas realizações, a distância entre a câmera e o ponto de intersecção associado a um objeto pode ser determinada pela estação de trabalho de processamento de conteúdo 10. Conseqüentemente, em algumas realizações, a quantidade de deslocamento pode ser gerada na dependência da distância entre a câmera e o ponto de intersecção para aquele objeto e aplicada como a quantidade de deslocamento para a totalidade daquele objeto.

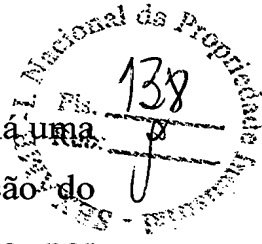
Em outras palavras, um jogador apareceria como bidimensional, porém apareceria como se estivesse posicionado em três dimensões no campo de futebol, a uma distância predeterminada do visor. Isto reduz vantajosamente recursos de processamento pois a distância a cada ponto em um jogador correspondendo a um pixel de saída no visor, não tem que ser detectada e usada para gerar uma respectiva quantidade de deslocamento. Ainda mais, isto equaciona um problema de que tais dados de distância podem não estar disponíveis a partir dos dados de rastreamento e posição gerados conforme descrito acima.

Em algumas realizações, a estação de trabalho de processamento de conteúdo 10 é operável para mapear um modelo tridimensional de um estádio, compreendendo o campo de futebol 30 para a imagem capturada, de tal modo que a estação de trabalho de processamento de conteúdo 10 pode gerar uma quantidade de deslocamento apropriada para cada pixel na imagem capturada correspondente ao estádio, de modo a fazer com que o estádio e/ou campo 30 apareçam como uma imagem tridimensional quando visualizados no visor 705. Como o estádio e o campo são relativamente estáticos com respeito à câmera 20, a geração das respectivas quantidades de deslocamento para cada pixel na imagem capturada pode ser realizada quando a imagem de segundo plano é gerada, ou pode ser realizada periodicamente, de modo a reduzir recursos de processamento.

No sentido de reduzir a probabilidade de que artefatos de imagem indesejáveis possam ocorrer na imagem modificada, quando a imagem direita (imagem modificada) é combinada com a imagem de segundo plano, em algumas realizações, a estação de trabalho de processamento de conteúdo 10 é operável para gerar uma imagem de segundo plano do campo 30, conforme descrito acima com referência à Figura 2 para cada quadro capturado. Isto permite ajuste da imagem de segundo plano de acordo com qualquer mudança na iluminação ou sombras no campo 30. Entretanto, será verificado que a imagem de segundo plano pode ser gerada e atualizada e qualquer outro intervalo de quadro adequado, por exemplo, a cada outro quadro.

A estação de trabalho é operável para mapear um modelo tridimensional do campo para a imagem capturada e gerar uma quantidade de deslocamento apropriada para cada pixel correspondente ao campo, conforme descrito acima, de modo a gerar uma imagem de segundo plano modificada. A estação de trabalho pode então combinar a imagem direita modificada correspondendo a um objeto tal como um jogador com a imagem de segundo plano modificada, de modo a gerar uma imagem modificada combinada. Por exemplo, a estação de trabalho 10 pode gerar a imagem modificada combinada, superpondo a imagem modificada correspondente a um objeto na imagem de segundo plano modificada. Quando a imagem capturada e a imagem modificada combinada são visualizadas juntas em um visor, de uma forma adequada, aparecerão para o usuário como se fossem uma imagem tridimensional.

Conforme mencionado acima, os dados de rastreamento indicam onde, no campo, está um jogador particular. Isto permite que a estação de trabalho de processamento de conteúdo 10 gere uma quantidade de deslocamento para aquele jogador, de modo a fazer com que apareça em uma posição tridimensional no campo, quando as imagens esquerda e direita são



visualizadas juntas no visor, como um par de imagens. Entretanto, se há uma ambigüidade pela qual a trilha pertence a qual jogador, ou a posição do jogador não pode ser determinada pelo algoritmo de rastreamento por qualquer que seja a razão, a aparência 3D simulada pode ser perdida ou degradada.

Conseqüentemente, em algumas realizações da invenção, o sistema compreende um detector de distância 810. O detector de distância 810 pode ser acoplado à câmera 20 ou pode ser separado da câmera 20. O detector de distância é operável para gerar dados de distância indicativos da distância entre a câmera 20 e um objeto tal como um jogador no campo 30. O detector de distância 810 é operável para enviar os dados de distância à estação de trabalho de processamento de conteúdo 10, via enlace de comunicação adequado, conforme indicado pela linha tracejada 812 na Figura 8. A estação de trabalho de processamento de conteúdo 10 é então operável para determinar a distância entre a câmera e o objeto, na dependência dos dados de distância recebidos do detector de distância 810. Em outras palavras, o detector de distância 810 atua como um sensor de distância. Tais sensores são conhecidos na técnica e podem ser de luz infravermelha, ultra-som, luz de laser e similares, para detectar distância a objetos.

Em algumas realizações, o detector de distância é operável para gerar dados de um mapa de distância que indica, para cada pixel da imagem capturada, uma respectiva distância entre a câmera e uma característica de cena dentro da cena, que coincide com aquele pixel. Os dados de distância enviados a partir do detector de distância 810 para a estação de trabalho 10 pode então compreender os dados do mapa de distância.

Para alcançar esta funcionalidade, o detector de distância pode compreender uma fonte luz infravermelha que indica impulso de luz infravermelha. A câmera 20 pode então detectar a intensidade da luz

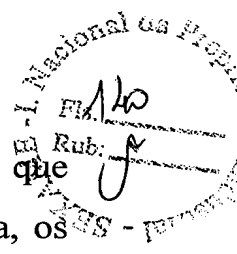
infravermelha detectada a partir de objetos dentro do campo de visão da câmera, a intervalos de tempo predeterminados (tipicamente da ordem de nano segundos) de modo a gerar uma imagem em escala de cinza indicativa da distância de objetos da câmera. Em outras palavras, a imagem em escala de cinza pode ser considerada como um mapa de distância que é gerado a partir da detecção do tempo de percurso da luz infravermelha da fonte para a câmera.

Para simplificar o projeto, a câmera pode compreender um detector de distância na forma de uma fonte luz infravermelha. Tais câmeras são conhecidas da técnica como a “Z-Cam” fabricada por 3DV Systems. Entretanto, será verificado que outros métodos conhecidos para gerar mapas de profundidade 3D poderiam ser usados, tais como detecção de distorção de configuração de infravermelho.

Será verificado que qualquer outro detector de distância adequado poderia ser usado. Por exemplo, uma câmera possuindo um eixo óptico que é perpendicular ao eixo óptico da câmera 20 pode ser usado para capturar imagens do campo. Estas imagens capturadas adicionais podem ser analisadas pela estação de trabalho 10 para detectar e rastrear as posições do jogador e os dados resultantes correlacionados ao dados de imagem a partir da câmera 20, de modo a triangular a posição dos jogadores mais precisamente.

Em algumas realizações, a estação de trabalho 10 é operável para usar o detector de distância 810 para detectar e rastrear outros objetos no campo de visão de câmera 20, tal como futebol, embora seja verificado que qualquer outro objeto poderia ser detectado. Por exemplo, imagens capturadas por uma ou mais câmeras adicionais podem ser analisadas pela estação de trabalho 10 e combinadas com dados a partir do sistema de rastreamento, de modo a rastrear o futebol e gerar conseqüentemente imagens esquerda e direita apropriadas.

Para gerar a imagem modificada, a estação de trabalho 10 é



operável para detectar pixéis de objeto dentro da imagem capturada, que correspondem ao objeto dentro da cena. Nas realizações descritas acima, os pixéis de objeto correspondem a aqueles pixéis da máscara do jogador usados para gerar a imagem modificada, conforme descrito acima. A estação de trabalho pode então determinar a distância entre a câmera e o jogador, usando os dados de distância que estão associados aos pixéis da máscara do jogador nos dados do mapa de distância. Para simplificar a visualização tridimensional uma média de valores de distância nos dados do mapa de distância que corresponde aos pixéis da máscara do jogador, pode ser usada para gerar a quantidade de deslocamento conforme descrito acima. Entretanto, será verificado que qualquer outro método adequado para selecionar um valor de distância a partir dos dados do mapa de distância correspondentes a um objeto, poderia ser usado.

Em outras realizações, a estação de trabalho é operável para gerar uma quantidade de deslocamento para aplicar entre a imagem esquerda e a imagem direita para cada pixel nos dados de mapa de profundidade. Conseqüentemente, quando a imagem esquerda e a imagem direita são visualizadas juntas como um par de imagens no visor, conforme descrito acima, os objetos podem ter uma aparência tridimensional melhorada porque a dimensão de superfície dos objetos pode ser reproduzida mais precisamente, ao invés de exibir o objeto como se fosse uma imagem bidimensional a alguma distância do visor.

Entretanto, através de distâncias de dezenas de metros, que são tipicamente envolvidas ao filmar esportes ou quadras de esportes, tais sistemas de câmera 3D podem perder a resolução de profundidade necessária para serem capazes de gerar uma quantidade de deslocamento para cada pixel correspondente a um objeto detectado dentro de uma imagem capturada, de modo a apresentar uma aparência tridimensional efetiva para um usuário. Portanto, as realizações descritas acima nas quais a quantidade de

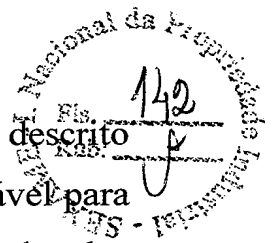


deslocamento é gerada na dependência dos dados de distância, nos dados de mapa de profundidade, são mais aplicáveis a situações nas quais objetos relevantes dentro de uma cena estão dentro de poucos metros da câmera. No contexto do material não editado de esportes, tais esportes podem ser boxe, sinuca, tênis de mesa, “jogo da pulga”, ginástica, esgrima e similares, embora seja verificado que submetidos a resolução de profundidade suficiente, outros eventos ou cenas poderiam também ser capturados.

Em algumas realizações, os dados de distância a partir do mapa de profundidade podem ser combinados com os dados de rastreamento e posição, de modo a melhorar a precisão ao determinar a posição do jogador no campo e daí a distância entre a câmera e o jogador.

Quando as realizações acima descritas são usadas para capturar imagens de jogadores em um campo de futebol, um possível problema pode surgir, se todos os jogadores estão dentro da mesma área no campo, por exemplo, durante um escanteio, tiro livre ou durante um ataque. Neste caso, quando uma imagem 3D é reproduzida no visor, a aparência 3D para um usuário pode se tornar confusa ou não clara, devido à oclusão de um jogador por outro jogador.

Conseqüentemente, nas realizações da presente invenção, a estação de trabalho de processamento de conteúdo 10 é operável para detectar se quaisquer dos jogadores estão dentro de uma distância limite um do outro, dentro da imagem capturada. Aqueles jogadores dentro da distância limite um do outro são então designados como pertencendo a um grupo de jogadores. Usando técnicas conforme descrito acima, a estação de trabalho é operável para determinar a respectiva distância entre a câmera e cada um dos jogadores, de modo a gerar um valor de distância médio indicativo da distância média entre a câmera e cada um dos jogadores no grupo. A estação de trabalho de processamento de conteúdo 10 é então operável para gerar uma quantidade de deslocamento, aplicável ao jogadores dentro do grupo, na



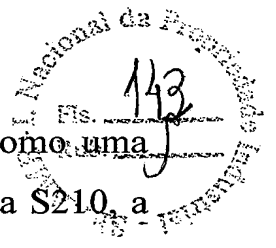
dependência do valor da distância média, de um modo similar ao descrito acima com referência às Figuras 8 a 10. A estrutura 10 é então operável para gerar a imagem direita deslocando, com respeito às posições detectadas dos jogadores dentro do grupo, cada jogador da quantidade de deslocamento para aquele grupo.

Isto é ilustrado na Figura 11, que é um diagrama de um grupo de jogadores na quadra. Em particular, Figura 11 ilustra os jogadores 1001, 1003 e 1005 no campo 30. Conforme mencionado acima, a estação de trabalho de processamento de conteúdo 10 é operável para detectar se os jogadores estão dentro de uma distância limite um do outro. No exemplo mostrado na Figura 11, os jogadores 1001, 1003 e 1005 estão dentro da distância limite um do outro e são então designados como pertencendo a um grupo de jogadores (conforme indicado pela linha tracejada 1007).

A estação de trabalho 10 é operável para determinar a distância entre cada um dos jogadores no grupo 1007, conforme descrito acima e gerar um valor de distância média para os jogadores naquele grupo. A quantidade de deslocamento é então gerada pela estação de trabalho de processamento de conteúdo 10 na dependência do valor de distância média e aplicada como a quantidade de deslocamento de imagem  $i$  para cada jogador 1001, 1003 e 1005 no grupo 1007. Quando as imagens esquerda e direita resultante são exibidas juntos como uma imagem 3D, todos os três jogadores no grupo aparecerão então na mesma distância do visor 705. Isto simplifica a visualização 3D, reduzindo então a fadiga do olho para um usuário, especialmente se há uma grande quantidade de ação rápida de jogo como ocorre freqüentemente durante um jogo de futebol.

Um método de processamento de imagem de acordo com realizações da presente invenção será agora descrito com referência à Figura 12.

Em uma etapa S200, a estação de trabalho de processamento



de conteúdo 10 recebe uma imagem capturada de uma cena, tal como uma  
imagem de um jogo de futebol, da câmera 20. Então, em uma etapa S210, a  
estação de trabalho de processamento de conteúdo 10 determina a distância  
entre a câmera e um objeto dentro da cena, tal como o jogador. A  
5 determinação pode ser baseada na posição e dados de rastreamento, gerada a  
partir dos dados de distância, gerados a partir dos dados do mapa de  
profundidade, ou a partir de uma combinação de alguns ou todos estes dados,  
usando as técnicas descritas acima.

Em uma etapa S220, a estação de trabalho 10 detecta uma  
10 posição do objeto dentro da imagem capturada usando as técnicas descritas  
acima. Por exemplo, em realizações, a estação de trabalho pode gerar uma  
máscara de jogador para cada jogador e usar as máscaras de jogador ao gerar  
a imagem modificada (imagem direita).

Então, em uma etapa S230, a estação de trabalho 10 gera a  
15 imagem modificada (imagem direita) deslocando a posição do objeto dentro  
da imagem modificada, da quantidade de deslocamento. Quando a imagem de  
captura (imagem esquerda) e a imagem modificada (imagem direita) são  
visualizadas juntas como um par de imagens em um visor (por exemplo,  
conforme mostrado na Figura 10), o objeto aparecerá posicionado a uma  
20 distância predeterminada do visor. Isto permite vantajosamente que imagens  
assim chamadas 3D sejam geradas sem a necessidade de pares de câmeras  
estéreo dispendiosas ou consumindo tempo em operações de pós  
processamento.

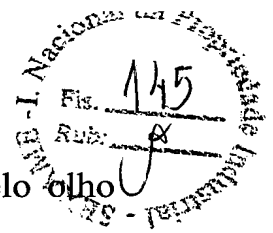
Embora as realizações acima tenham sido descritas com  
25 referência a jogadores de um jogo de esportes, será verificado que as técnicas  
acima descritas podem ser aplicadas a outras situações, onde é desejável  
capturar e produzir imagens 3D.

Ainda mais, será verificado que, embora as realizações acima  
descritas tenham sido descritas como realizando processamento de imagem

em uma imagem, será verificado que as técnicas acima podem ser aplicadas a imagens em uma seqüência de imagens de vídeo capturadas por uma câmera de vídeo. Deslocando a posição de um objeto dentro de uma imagem modificada com respeito a uma posição detectada do objeto dentro da imagem capturada, de uma quantidade de deslocamento que é dependente de uma distância determinada pela estação de trabalho a partir de dados tais como dados de imagem ou dados de distância, imagens 3D podem ser capturadas e geradas de tal modo que podem ser visualizadas por um usuário, substancialmente em tempo real.

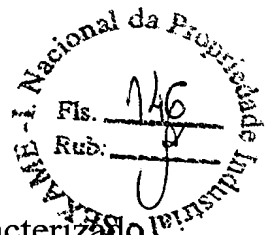
Nas realizações descritas acima, processamento de imagem foi descrito como sendo realizado sobre uma imagem capturada pela câmera 20. Entretanto, será verificado que as técnicas acima descritas são igualmente aplicáveis a imagens de vídeo que tenham sido “costuradas” juntas, por exemplo, de modo a formar imagens de ultra alta resolução tais como aquelas geradas “costurando” juntas imagens de vídeo capturadas pelas câmeras 22.1 e 22.2. Adicionalmente, será verificado que a imagem capturada poderia ser tratada como a imagem direita e a imagem modificada como a imagem esquerda.

Será verificado que no sentido de obter uma aparência tridimensional para um usuário, a imagem esquerda (L) e a imagem direita (R) na necessitam ser exibidas simultaneamente em uma tela. Por exemplo, estas podem ser exibidas alternadamente a uma taxa que corresponde a uma taxa de quadro para cada imagem, que é mais rápida que a visão persistente de um usuário (tipicamente 24 quadros por segundo). Em outras palavras, uma seqüência de quadros compreendendo imagens esquerda e direita alternadas (por exemplo, LRLRLRLRLR) poderiam ser exibidas a uma taxa de quadro de 60 quadros por segundo (isto é, 30 quadros por segundo para cada imagem), embora qualquer outra taxa de quadro adequada possa ser usada. O usuário pode então visualizar as imagens usando um par apropriado de óculos que apagam alternadamente a imagem esquerda ou imagem direita



correspondente, de tal modo que a imagem correta é visualizada pelo olho  
correto. Portanto, o termo “visualizadas juntas” deveria ser entendido como  
significando que a imagem esquerda e a imagem direita podem ser  
visualizadas simultaneamente, que podem ser visualizadas alternadamente ou  
5 que podem ser visualizadas de qualquer outro modo adequado, de tal modo  
que o usuário perceba um efeito tridimensional.

Será verificado que, nas realizações da presente invenção,  
elementos do método de processamento de imagem podem ser implementados  
na estação de trabalho de processamento de conteúdo de qualquer maneira  
10 adequada. Então, a adaptação requerida para partes existentes de um  
dispositivo equivalente convencional pode ser implementada na forma de um  
produto de programa de computador compreendendo instruções de  
processador implementáveis armazenadas em um portador de dados tal como  
um disco flexível, disco óptico, disco rígido, PROM, RAM, memória flash ou  
15 qualquer combinação destes ou de outros meios de armazenagem, ou  
transmitidas via sinais de dados em uma rede tal como uma Ethernet, uma  
rede sem fio, Internet ou qualquer combinação destas ou outras redes, ou  
realizada em hardware como um ASIC (circuito integrado específico da  
aplicação) ou um FPGA (arranjo de porta programável no campo) ou outro  
20 circuito configurável ou ajustável, adequado para uso na adaptação do  
dispositivo equivalente convencional.



## REIVINDICAÇÕES

1. Dispositivo de processamento de imagem, caracterizado pelo fato de compreender:

meio de recepção operável para receber, de uma câmera, uma  
5 imagem capturada correspondente a uma imagem de uma cena capturada pela câmera, a cena contendo pelo menos um objeto;

meio de determinação operável para determinar uma distância entre o objeto dentro da cena e uma posição de referência definida com respeito à câmera; e

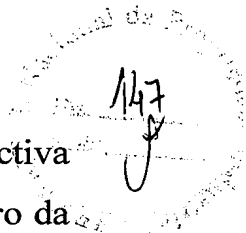
10 meio de geração operável para detectar uma posição do objeto dentro da imagem capturada, e para gerar uma imagem modificada a partir da imagem capturada, com base nas características de imagem dentro da imagem capturada que correspondem ao objeto na cena,

em que:

15 o meio de geração é operável para gerar a imagem modificada, deslocando a posição do objeto capturado dentro da imagem modificada com respeito à posição determinada do objeto dentro da imagem capturada por uma quantidade de deslocamento que é dependente da distância entre a posição de referência e o objeto na cena, de tal modo que, quando a imagem  
20 modificada e a imagem capturada sejam visualizadas juntas como um par de imagens em um visor, o objeto capturado aparece como posicionado a uma distância predeterminada do visor.

2. Dispositivo de acordo com a reivindicação 1, caracterizado pelo fato de que o meio de determinação é operável para determinar a  
25 distância entre a posição de referência e o objeto na cena, na dependência dos dados de distância, os dados de distância sendo indicativos da distância entre a posição de referência e o objeto na cena.

3. Dispositivo de acordo com a reivindicação 2, caracterizado pelo fato de que os dados de distância compreendem dados de mapa de



distância que indicam, para cada pixel da imagem capturada, uma respectiva distância entre a posição de referência e uma característica de cena dentro da cena coincidindo com aquele pixel.

4. Dispositivo de acordo com a reivindicação 3, caracterizado pelo fato de que o meio de geração é operável para detectar pixéis de objeto dentro da imagem capturada que correspondem ao objeto dentro da cena.

5. Dispositivo de acordo com a reivindicação 4, caracterizado pelo fato de que:

os pixéis de objeto formam juntos uma imagem de primeiro plano, e aqueles pixéis que são detectados como não correspondentes ao objeto dentro da cena são designados como pixéis de segundo plano que formam uma imagem de segundo plano;

o meio de geração é operável para deslocar a posição dos pixéis de objeto com respeito à posição do objeto dentro da imagem capturada pela quantidade de deslocamento, de modo a gerar uma imagem de primeiro plano modificada;

o meio de determinação é operável para determinar a distância entre a posição de referência e características de cena na cena, que coincidem com os respectivos pixéis de segundo plano na dependência dos dados de distância associados ao mapa de distância com aqueles pixéis de segundo plano;

o meio de geração é operável para gerar a imagem modificada, deslocando a posição de cada pixel de segundo plano com respeito à posição daquele pixel de segundo plano na imagem de segundo plano de uma respectiva quantidade de deslocamento de pixel de segundo plano, que é dependente da distância entre a posição de referência e a característica de cena dentro da cena, que corresponde a aquele pixel de segundo plano, de modo a gerar uma imagem de segundo plano modificada, e para combinar a imagem de segundo plano modificada com a imagem de primeiro plano

modificada, de tal modo que, quando a imagem modificada e a imagem capturada são visualizadas juntas como um par de imagens no visor, a imagem modificada e a imagem capturada aparecem juntas como tridimensionais.

5                   6. Dispositivo de acordo com a reivindicação 4, caracterizado pelo fato de que o meio de determinação é operável para determinar a distância entre a posição de referência e o objeto dentro da cena, na dependência dos dados de distância associados, nos dados de mapa de distância, a aqueles pixéis de objeto.

10                   7. Dispositivo de acordo com a reivindicação 4, caracterizado pelo fato de que o meio de geração é operável para gerar a imagem modificada deslocando a posição de cada pixel de objeto com respeito à posição daquele pixel na imagem capturada, por uma respectiva quantidade de desvio de pixel de objeto que é dependente da distância entre a posição de  
15                   referência e a característica de cena dentro da cena que corresponde a aquele pixel de objeto.

20                   8. Dispositivo de acordo com a reivindicação 2, caracterizado pelo fato de que um modelo tridimensional predeterminado de pelo menos parte da cena é mapeado para pontos de característica conhecidos dentro da cena; e

o meio de determinação é operável para determinar a distância entre a posição de referência e o objeto na cena, na dependência da posição detectada do objeto na cena, com respeito ao modelo tridimensional predeterminado.

25                   9. Dispositivo de acordo com a reivindicação 8, caracterizado pelo fato de que o meio de determinação é operável para detectar características de imagem predeterminadas dentro da imagem capturada, que correspondem a posições dos pontos de característica conhecidos dentro da cena, de modo a mapear o modelo tridimensional para os pontos de

característica conhecidos.

10. Dispositivo de acordo com a reivindicação 2, caracterizado pelo fato de que:

o meio de determinação é operável para detectar uma posição do objeto na cena, com respeito à câmera, e para determinar a distância entre o objeto na cena e a posição de referência, na dependência da posição detectada do objeto na cena, com respeito à câmera.

11. Dispositivo de acordo com a reivindicação 10, caracterizado pelo fato de que:

o meio de recepção é operável para receber uma seqüência de imagens capturadas pela câmera; e

o meio de determinação é operável para:

rastrear a posição do objeto capturado através da seqüência de imagens, na dependência da posição detectada do objeto na cena, com respeito à câmera, de modo a gerar dados de rastreamento; e

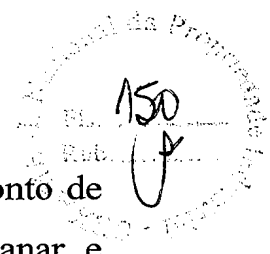
determinar a distância entre o objeto na cena e a posição de referência, na dependência dos dados de rastreamento.

12. Dispositivo de acordo com a reivindicação 2, caracterizado pelo fato de que o meio de determinação é operável para gerar os dados de distância pela análise da imagem capturada.

13. Dispositivo de acordo com a reivindicação 2, caracterizado pelo fato de compreender meio de recepção de dados de distância operável para receber os dados de distância de um sensor de distância, onde o meio de determinação é operável para determinar a distância entre a posição de referência e o objeto na cena, na dependência dos dados de distância recebidos do sensor de distância.

14. Dispositivo de acordo com a reivindicação 1, caracterizado pelo fato de que:

a cena compreende uma superfície substancialmente planar;



o meio de determinação é operável para detectar um ponto de interseção no qual o objeto dentro da cena coincide com a superfície planar, e para determinar a distância entre a posição de referência e o ponto de interseção; e

5 a quantidade de desvio de objeto é dependente da distância entre a posição de referência e o ponto de interseção.

15. Dispositivo de acordo com a reivindicação 1, caracterizado pelo fato de que o meio de geração é operável para detectar diversos objetos dentro da cena, pela análise da imagem capturada;

10 o meio de geração é operável para detectar se os diversos objetos dentro da cena estão dentro de uma distância limite um do outro, aqueles objetos que são detectados como estando dentro da distância limite um do outro sendo designados como pertencentes a um grupo de objetos dentro da cena;

15 o meio de determinação é operável para detectar uma respectiva distância entre a posição de referência e cada um dos objetos no grupo de objetos dentro da cena, de modo a gerar um valor de distância médio indicativo da distância média entre a posição de referência e cada um dos objetos no grupo de objetos dentro da cena;

20 e  
se os diversos objetos na cena são detectados como estando dentro da distância limite um do outro, o meio de geração é operável para gerar a imagem modificada, pelo deslocamento, com respeito a determinadas posições dos diversos objetos dentro da cena, de cada objeto capturado no grupo de objetos, por uma quantidade de deslocamento de grupo, sendo  
25 dependente do valor de distância média.

16. Dispositivo de acordo com a reivindicação 1, caracterizado pelo fato de que a posição de referência está localizada em uma posição da câmera.

17. Dispositivo de acordo com a reivindicação 1, caracterizado pelo fato de que:

a posição de referência permanece em uma linha de referência, a linha de referência sendo ortogonal a um eixo óptico da câmera e passando através de uma posição da câmera; e

a posição de referência está localizada na linha de referência em um ponto onde uma linha de localização de objeto e a linha de referência se interceptam, a linha de localização de objeto sendo ortogonal à linha de referência e passando pelo objeto.

18. Dispositivo de acordo com a reivindicação 17, caracterizado pelo fato de que a linha de referência é paralela a uma característica de cena dentro da cena.

19. Dispositivo de acordo com a reivindicação 1, caracterizado pelo fato de que:

o meio de recepção é operável para receber diversas imagens capturadas de diferentes câmeras respectivas, tendo eixos ópticos que se interceptam um com o outro, cada imagem capturada correspondendo a uma imagem diferente da cena; e

o meio de determinação é operável para determinar a distância entre o objeto dentro da cena e a posição de referência, pela análise das imagens capturadas recebidas das câmeras.

20. Dispositivo de acordo com a reivindicação 1, caracterizado pelo fato de que a cena compreende um estádio de esportes tendo um campo de esportes, e o objeto dentro da cena é um jogador no campo de esportes ou um objeto esportivo usado naquele esporte.

21. Sistema de processamento de imagem, caracterizado pelo fato de compreender:

uma ou mais câmeras operáveis para capturar uma imagem de uma cena, de modo a gerar uma imagem capturada; e



um dispositivo de processamento de imagem como definido na reivindicação 1.

22. Sistema de acordo com a reivindicação 21, caracterizado pelo fato de compreender um visor operável para exibir a imagem modificada juntamente com a imagem capturada, como um par de imagens no visor.

23. Método de processamento de imagem, caracterizado pelo fato de compreender:

receber, de uma câmera, uma imagem capturada correspondente a um imagem de uma cena capturada pela câmera, a cena compreendendo pelo menos um objeto;

determinar uma distância entre o objeto dentro da cena e uma posição de referência definida com respeito à câmera;

detectar uma posição do objeto dentro da imagem capturada; e

gerar uma imagem modificada a partir da imagem capturada, com base nas características de imagem dentro da imagem capturada que corresponde ao objeto na cena, deslocando a posição do objeto capturado dentro da imagem capturada, de uma quantidade de deslocamento de objeto que é dependente da distância entre a posição de referência e o objeto na cena, de tal modo que, quando a imagem modificada e a imagem capturada são visualizadas juntas como um par de imagens em um visor, o objeto capturado aparece como posicionado a uma distância predeterminada do visor.

24. Software de computador, caracterizado pelo fato de que, quando executado por um computador, faz com que o computador execute o método como definido na reivindicação 23.

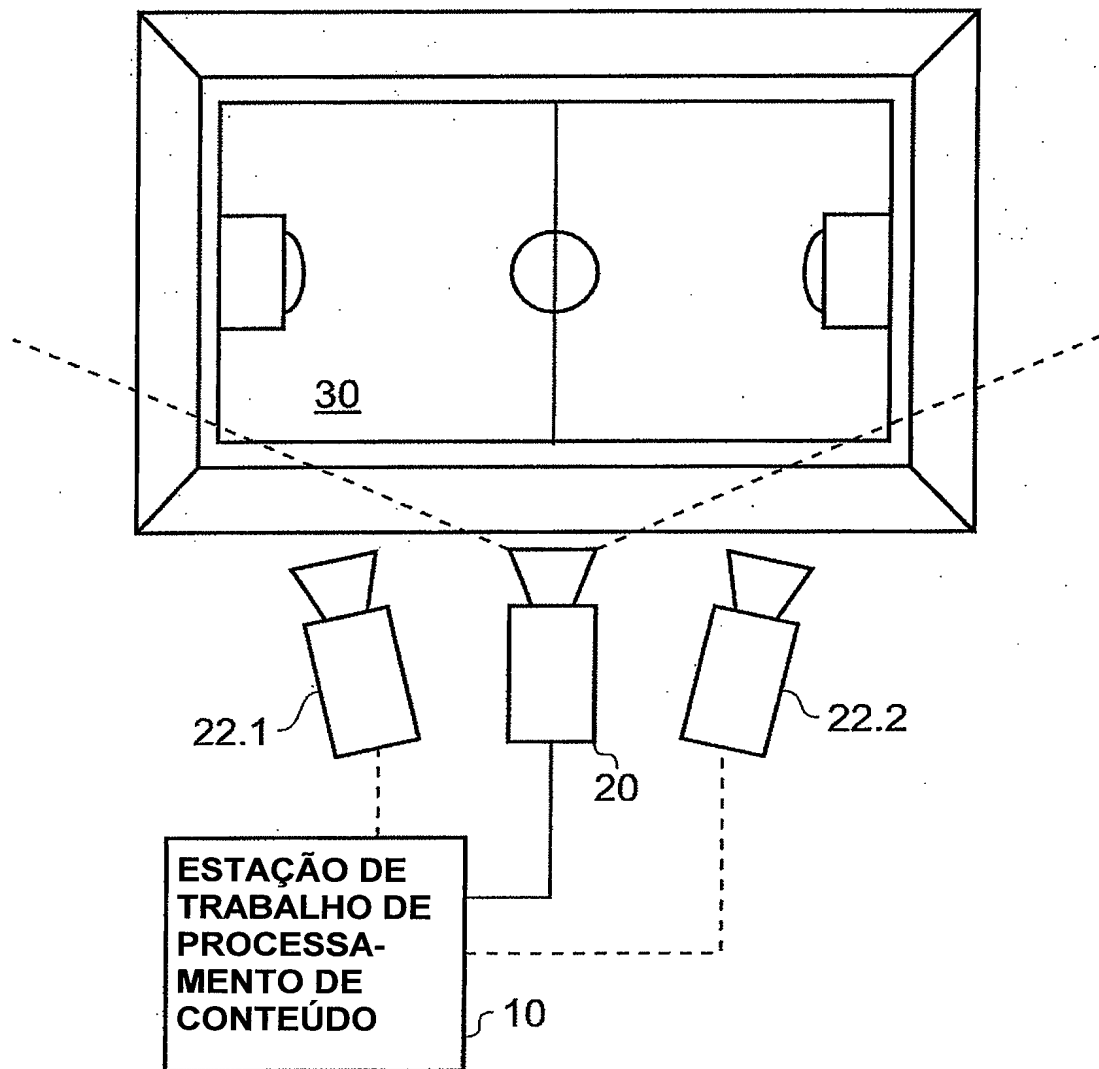


FIG. 1

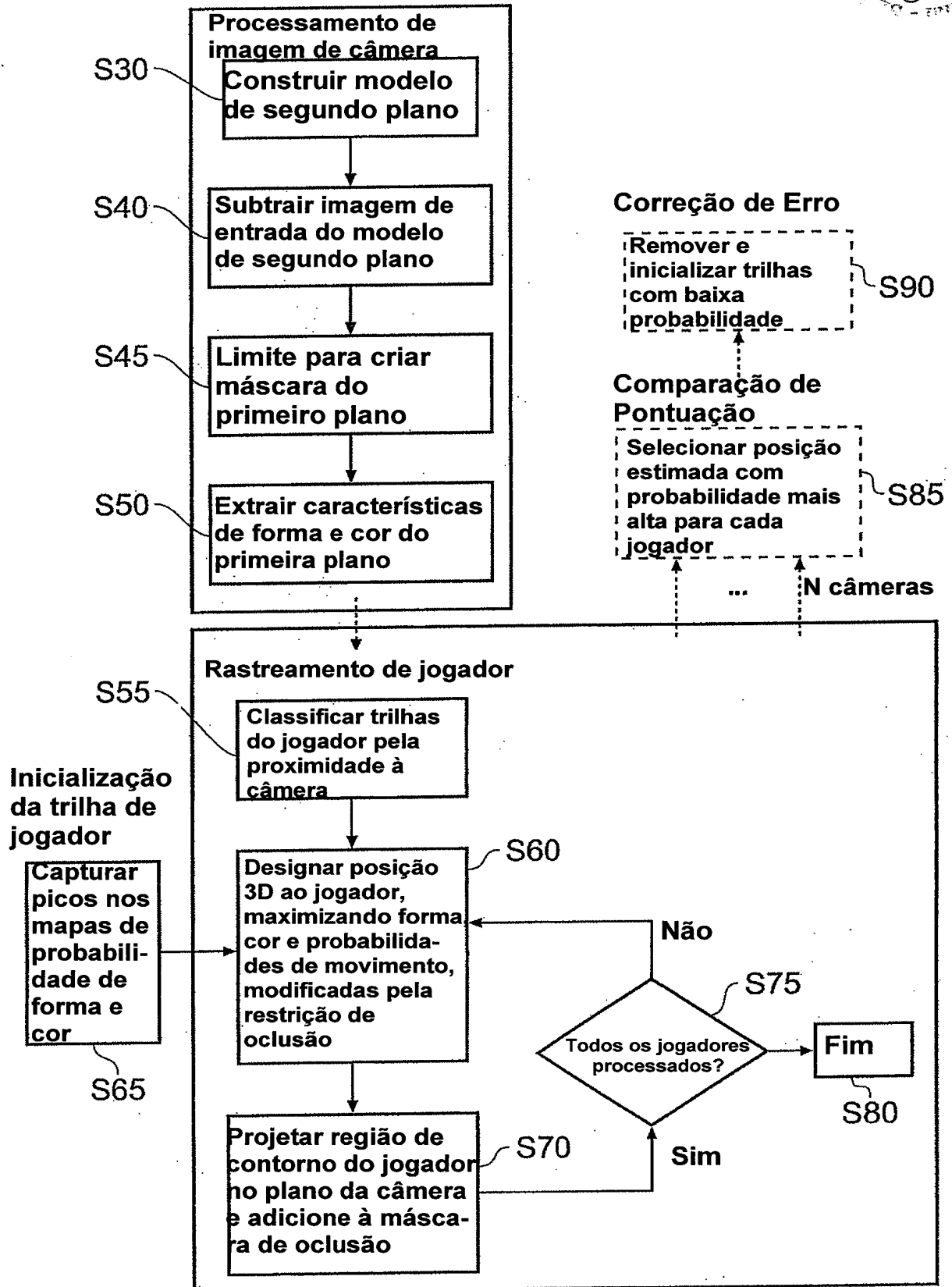


FIG. 2

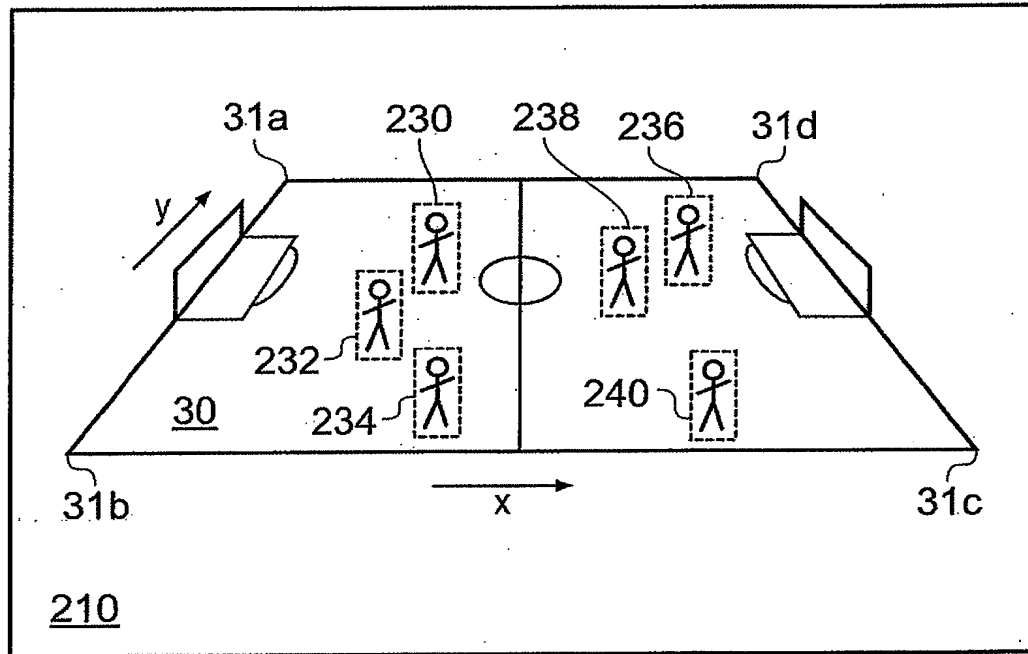


FIG. 3A

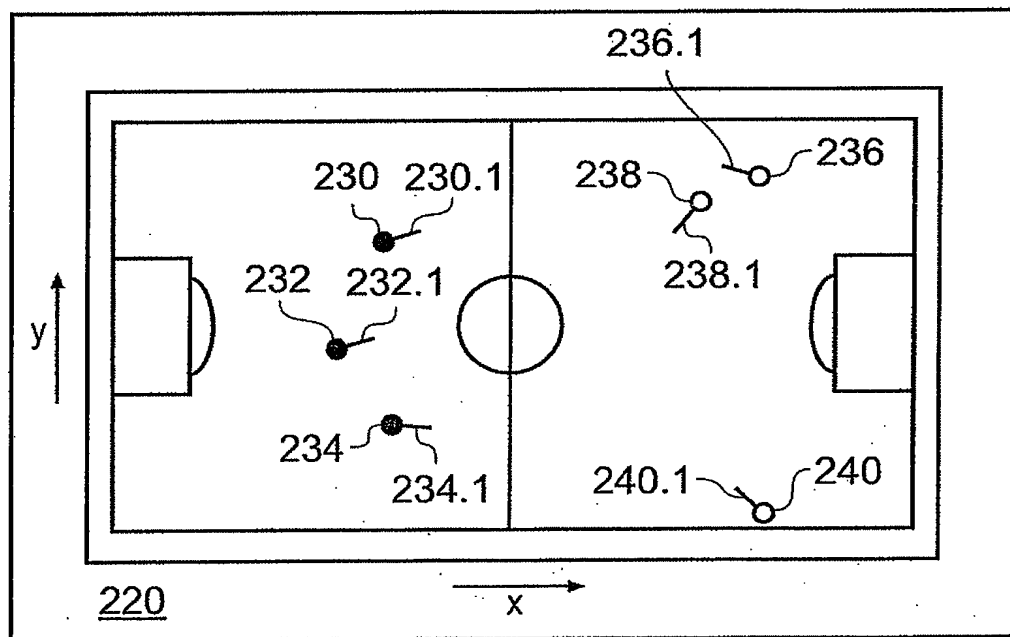


FIG. 3B

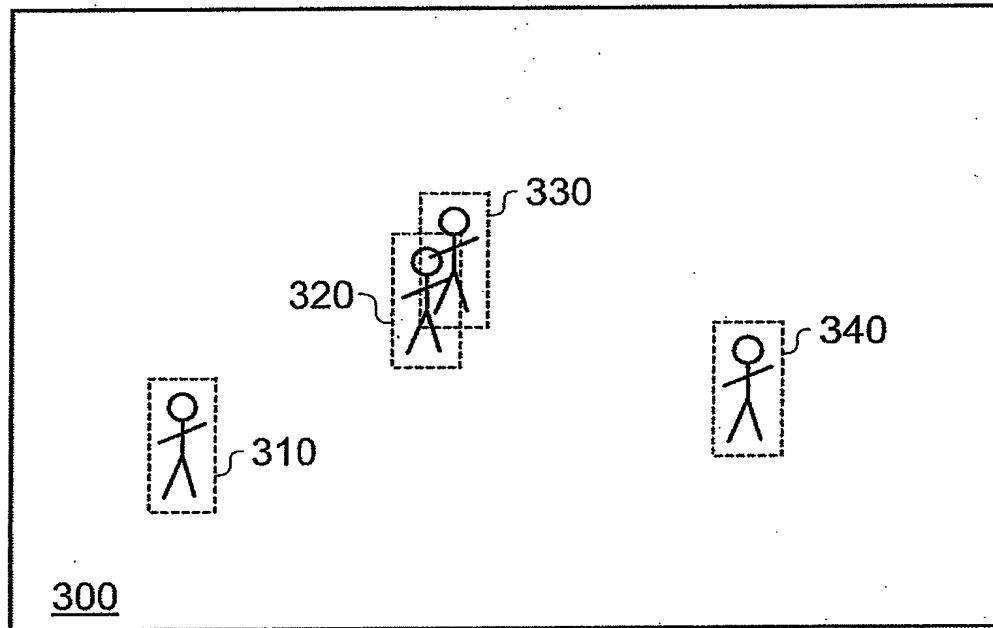


FIG. 4

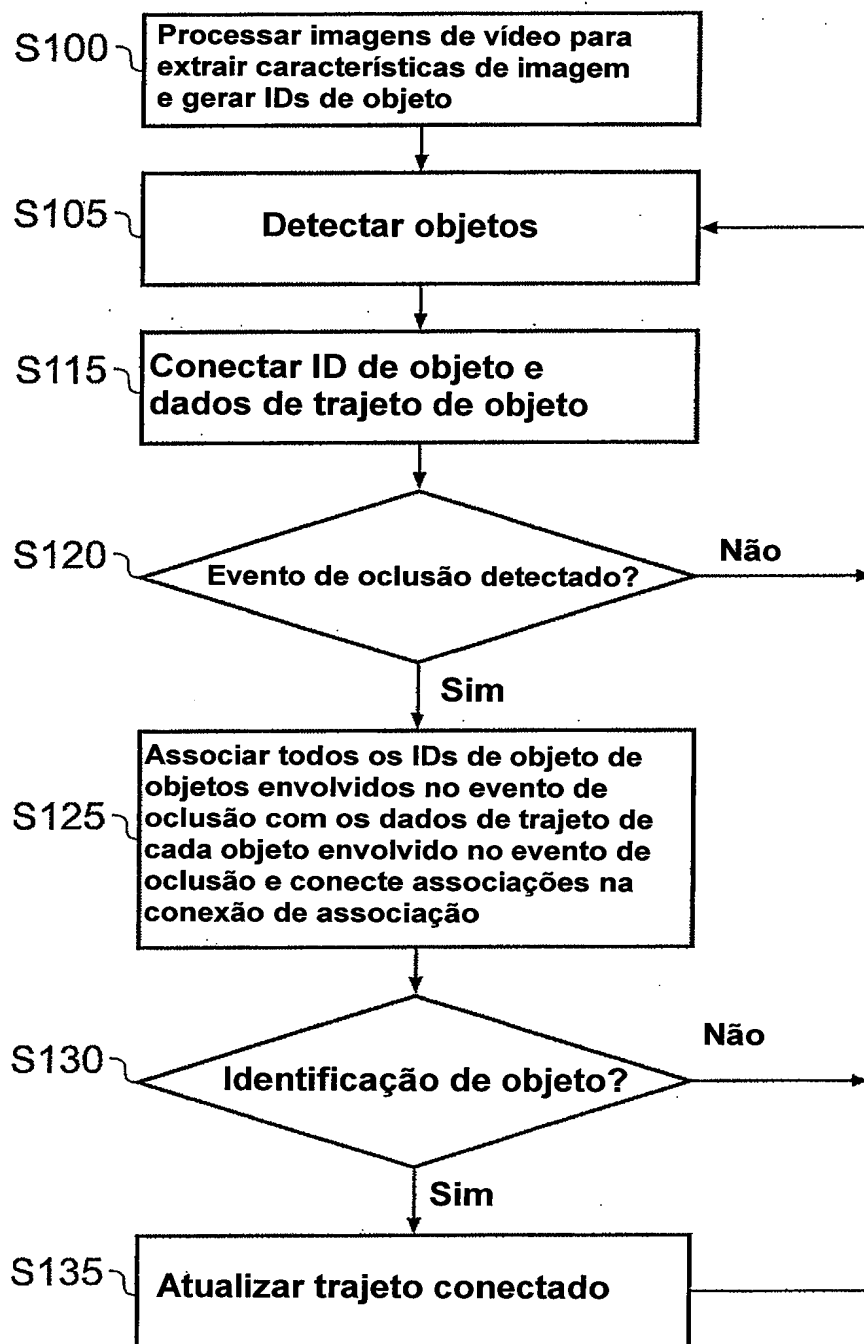


FIG. 5

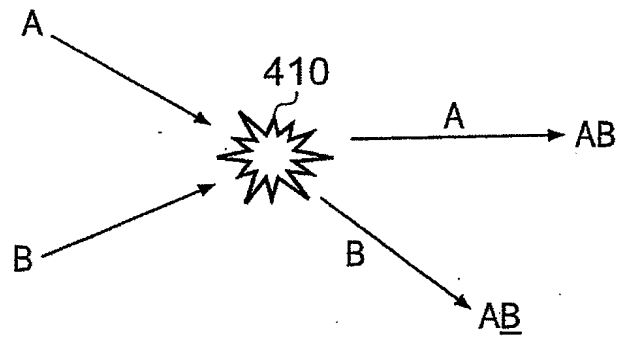


FIG. 6A

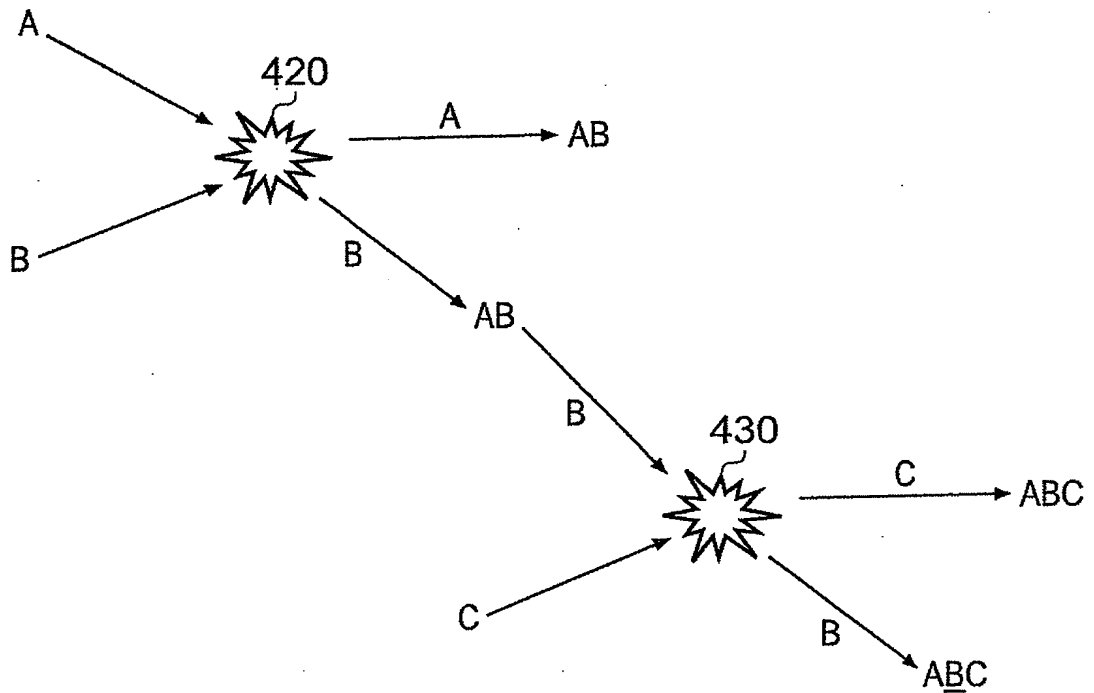


FIG. 6B

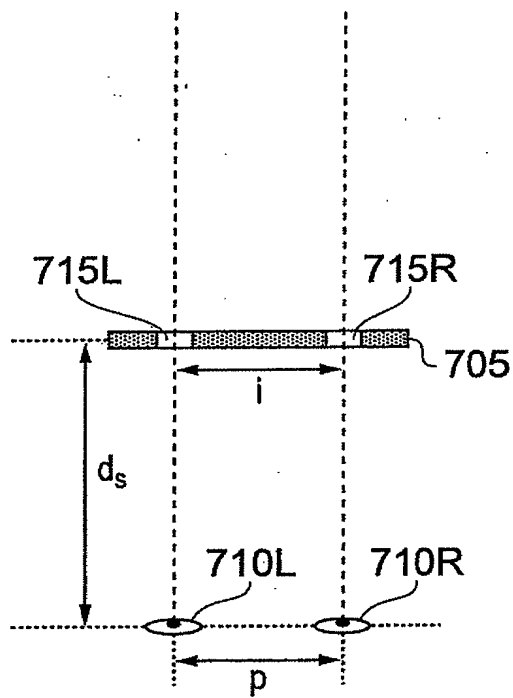


FIG. 7A

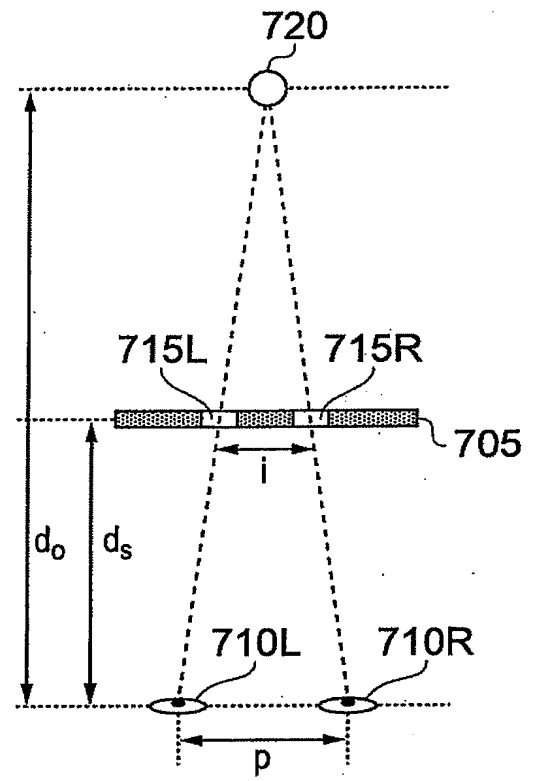


FIG. 7B

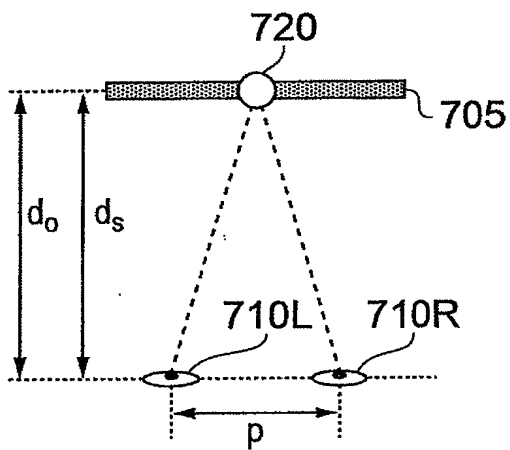


FIG. 7C

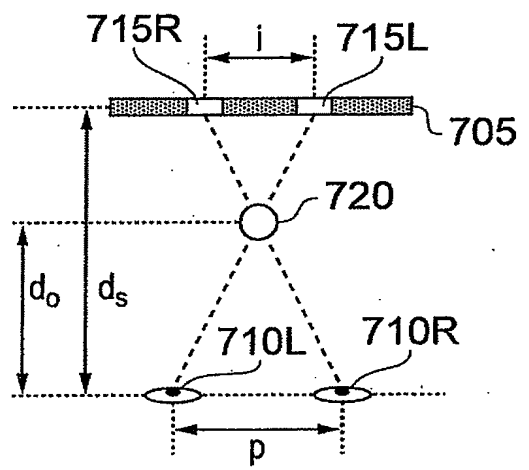


FIG. 7D

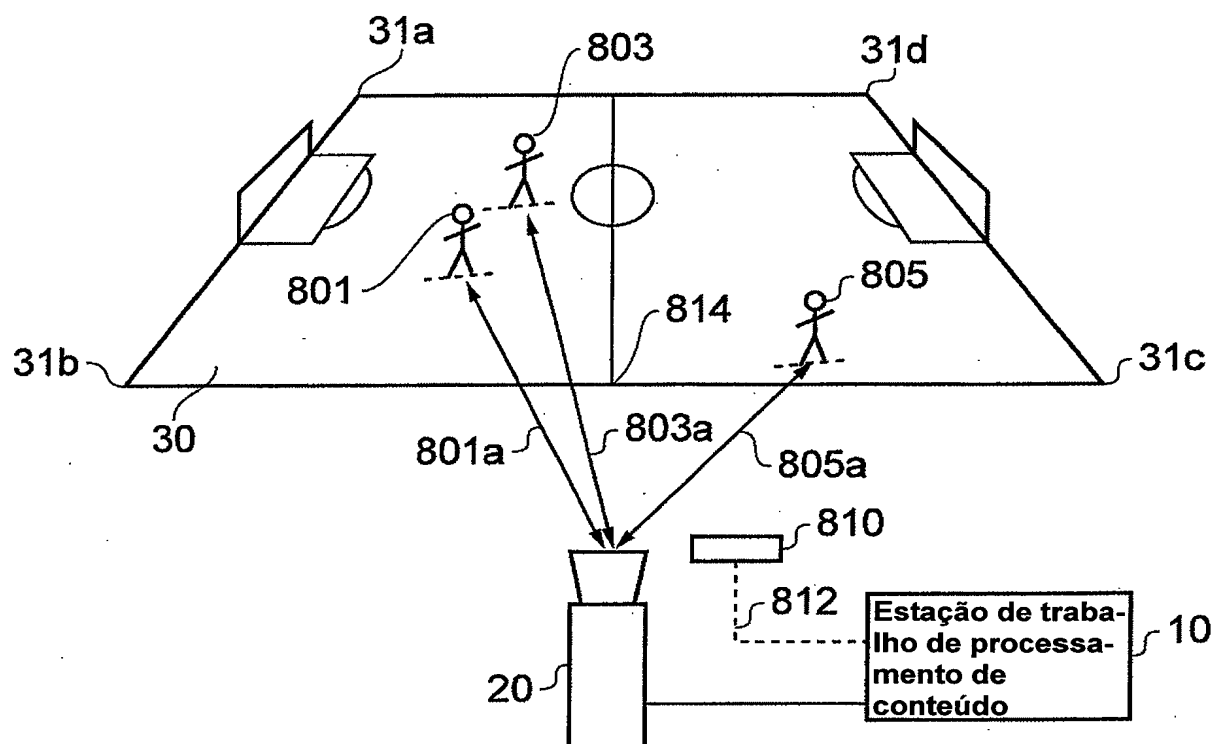


FIG. 8

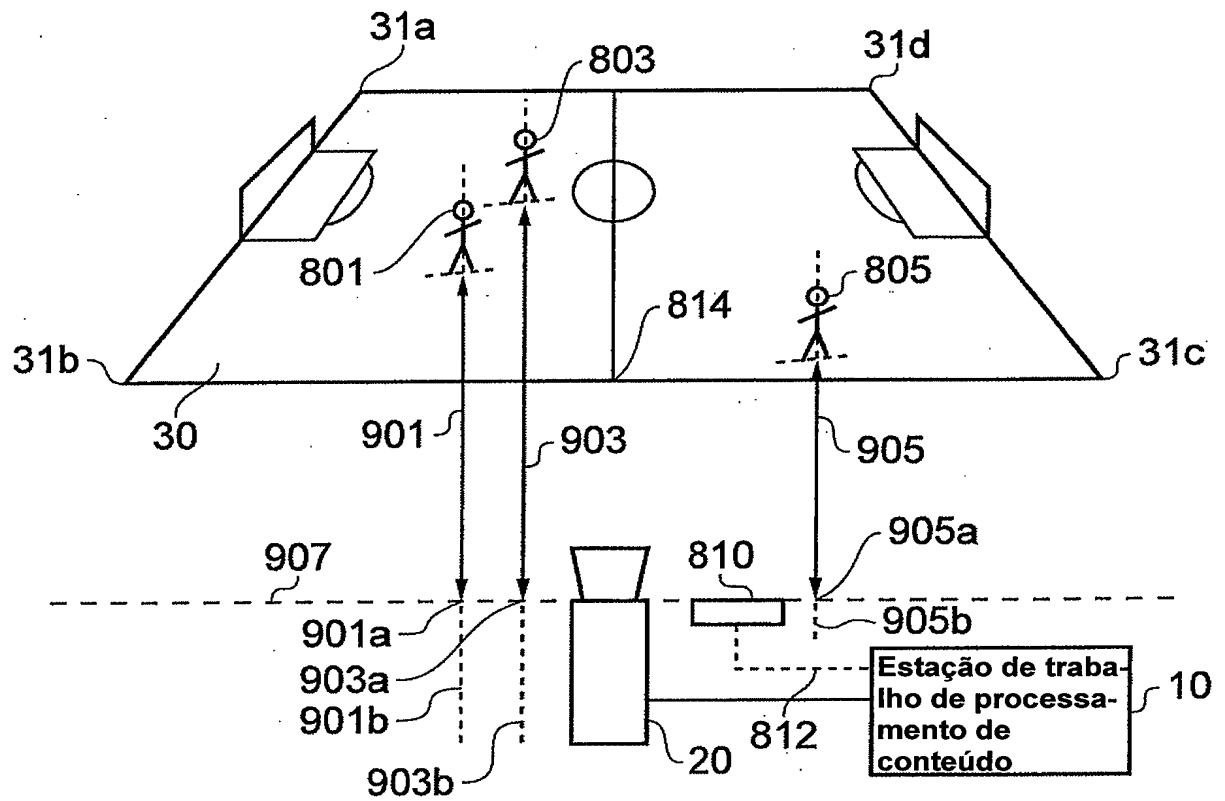
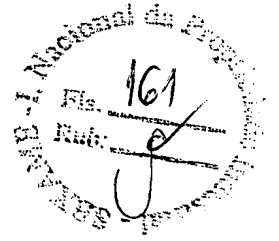


FIG. 9

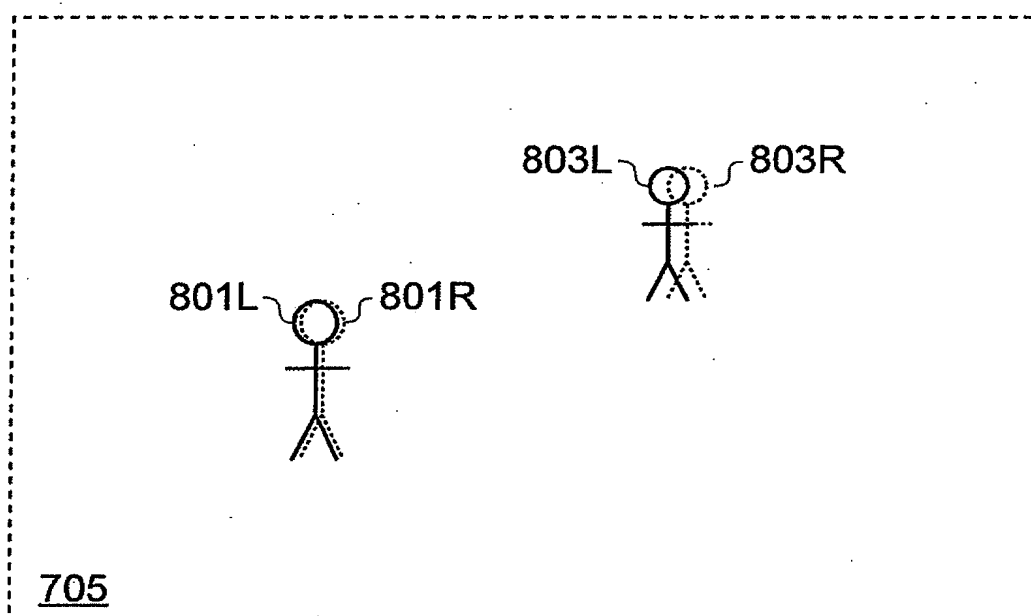


FIG. 10

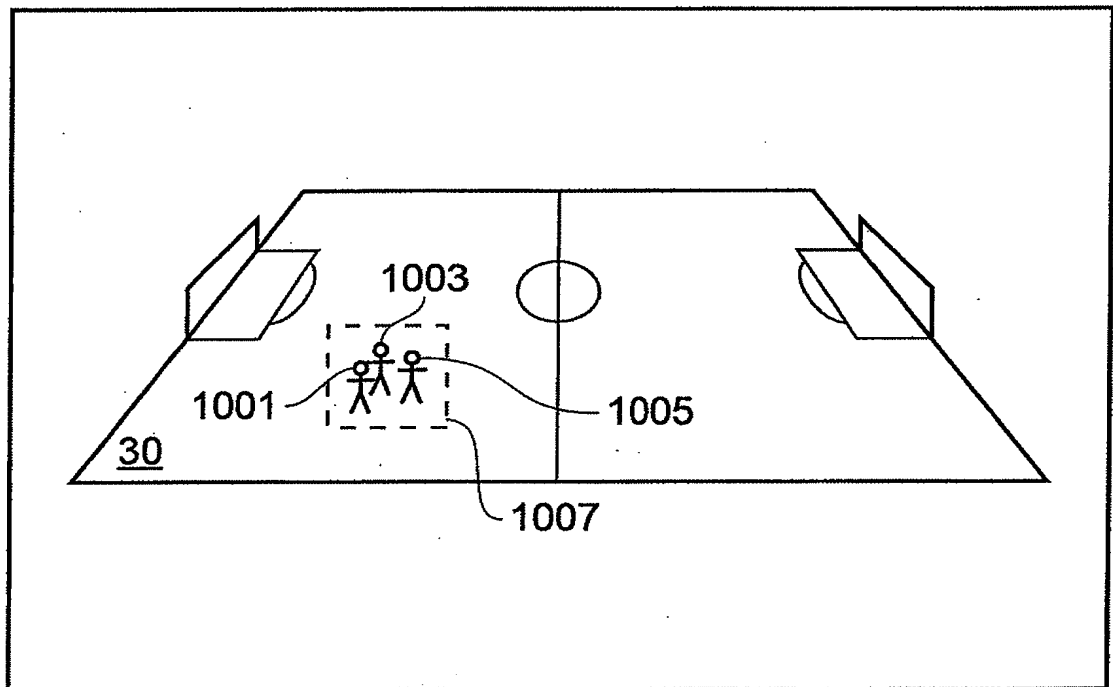
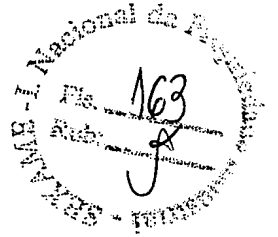


FIG. 11

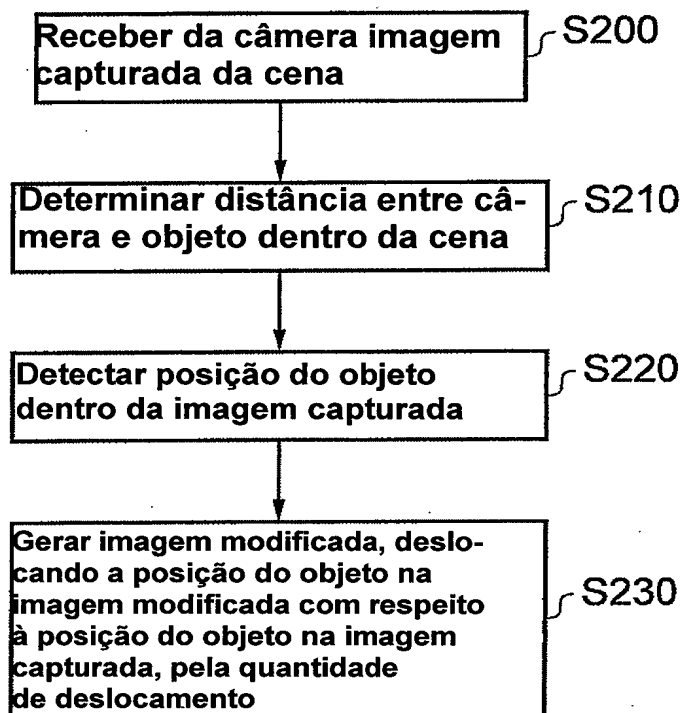
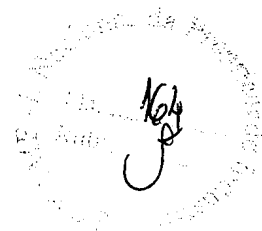
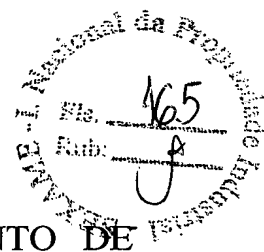


FIG. 12



## RESUMO

“DISPOSITIVO, SISTEMA E MÉTODO DE PROCESSAMENTO DE IMAGEM, E SOFTWARE DE COMPUTADOR”

Um dispositivo de processamento de imagem compreende meio operável para receber, de uma câmera, uma imagem capturada correspondente a uma imagem de uma cena capturada pela câmera. A cena contém pelo menos um objeto. O dispositivo compreende meio de determinação operável para determinar uma distância entre o objeto dentro da cena e uma posição de referência definida com respeito à câmera e meio de geração operável para detectar uma posição do objeto dentro da imagem capturada, e para gerar uma imagem modificada a partir da imagem capturada, com base nas características de imagem dentro da imagem capturada que correspondem ao objeto na cena. O meio de geração é operável para gerar a imagem modificada, deslocando a posição do objeto capturado dentro da imagem modificada com respeito à posição determinada do objeto dentro da imagem capturada por uma quantidade de deslocamento que é dependente da distância entre a posição de referência e o objeto na cena, de tal modo que, quando a imagem modificada e a imagem capturada são visualizadas juntas como um par de imagens em um visor, o objeto capturado aparece como posicionado a uma distância predeterminada do visor.