



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

(21) PI 0721452-9 A2



(22) Data de Depósito: 19/12/2007
(43) Data da Publicação: 25/03/2014
(RPI 2255)

(51) Int.Cl.:
H04N 13/00

(54) Título: SISTEMA E MÉTODO PARA COMBINAR TEXTO COM CONTEÚDO TRIDIMENSIONAL **(57) Resumo:**

(30) Prioridade Unionista: 16/03/2007 US 60/918.635

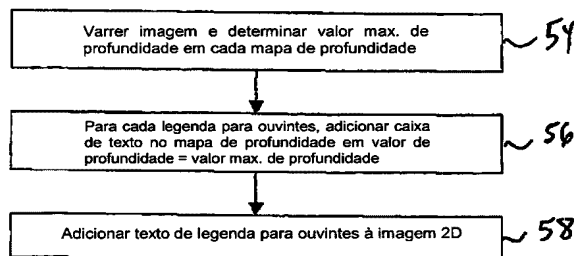
(73) Titular(es): Thomson Licensing

(72) Inventor(es): DONG-QING ZHANG, Izzat Izzat, Yousef Wasef Nijim

(74) Procurador(es): Nellie Anne Daniel-Shores

(86) Pedido Internacional: PCT US2007025947 de 19/12/2007

(87) Publicação Internacional: WO 2008/115222de 25/09/2008



“SISTEMA E MÉTODO PARA COMBINAR TEXTO COM CONTEÚDO TRIDIMENSIONAL”

Esse pedido reivindica o benefício de acordo com 35 U.S.C. § 119 de um pedido provisional 60/918635 depositado nos Estados Unidos em 16 de março de 2007.

Campo técnico da invenção

5 A presente revelação refere-se genericamente a sistemas de exibição e processamento de imagem, e mais particularmente, a um sistema e método para combinar texto com conteúdo tridimensional.

Antecedentes da invenção

10 Há dois tipos de texto que podem ser adicionados a vídeo: legendas para ouvintes e legendas para deficientes auditivos. Dito em termos gerais, legendas são destinadas a audiências que ouvem e legendas para audiências surdas. Legendas para ouvintes podem traduzir o diálogo em um idioma diferente, porém raramente mostram todo o áudio. Por exemplo, legendas para deficientes auditivos mostram efeitos de som (por exemplo, “telefone tocando” e “passos”), enquanto legendas para ouvintes não mostram.

15 Legendas ocultas são legendas que são ocultas em um sinal de vídeo, invisíveis sem um decodificador especial. As legendas ocultas são ocultas, por exemplo, na linha 21 do intervalo de supressão de linha (VBI). Legendas abertas são legendas que foram decodificadas, assim se tornaram parte integral da imagem da televisão, como legendas para ouvintes em um filme. Em outras palavras, legendas abertas não podem ser desligadas. O
20 termo “legendas abertas” também é utilizado para se referir a legendas para ouvintes criadas com um gerador de caracteres.

O uso de texto em vídeo bidimensional 2D é conhecido por aqueles versados na técnica. O interesse atual em filme e vídeo tridimensional 3D criou a necessidade de técnicas para acrescentar texto ao conteúdo 3D. Portanto, existe uma necessidade por técnicas
25 para otimizar a inserção de texto em conteúdo 3D de tal modo que o texto adicionado não obstrua os efeitos 3D no conteúdo 3D e não crie fadiga visual quando o conteúdo 3D é visto.

Sumário

De acordo com um aspecto da presente revelação, são fornecidos sistema e método para combinar e/ou exibir texto com conteúdo tridimensional 3D. O sistema e método
30 inserem texto no mesmo nível como o valor de profundidade mais elevado no conteúdo 3D. Um exemplo de conteúdo 3D é uma imagem bidimensional e um mapa de profundidade associado. Nesse caso, o valor de profundidade do texto inserido é ajustado para casar com o valor de profundidade maior do mapa de profundidade dado. Outro exemplo de conteúdo de 3D é uma pluralidade de imagens bidimensionais e mapas de profundidade associados.
35 Nesse caso, o valor de profundidade do texto inserido é continuamente ajustado para casar com o valor de profundidade maior de um mapa de profundidade dado. Um exemplo adicional de conteúdo 3D é conteúdo estereoscópico tendo uma imagem de olho direito e uma

imagem de olho esquerdo. Nesse caso o texto em uma da imagem de olho esquerdo e imagem de olho direito é deslocado para casar com o valor de profundidade maior na imagem estereoscópica. Ainda outro exemplo de conteúdo 3D é conteúdo estereoscópico tendo uma pluralidade de imagens de olho direito e imagens de olho esquerdo. Nesse caso o texto em uma das imagens de olho esquerdo ou imagens de olho direito é continuamente deslocada para casar com o valor de profundidade maior nas imagens estereoscópicas. Como resultado, o sistema e método da presente revelação produzem texto combinado com conteúdo 3D onde o texto não obstrui os efeitos 3D no conteúdo 3D e não cria fadiga visual quando visto por um telespectador.

De acordo com outro aspecto da presente revelação, um método para combinar texto com conteúdo de imagem tridimensional que recebe conteúdo de imagem tridimensional, determinar um valor de profundidade máximo para o conteúdo tridimensional, e combinar texto com conteúdo de imagem tridimensional no valor máximo de profundidade.

De acordo com um aspecto adicional da presente revelação, um método de exibir texto com conteúdo de imagem tridimensional inclui receber conteúdo de imagem tridimensional e texto, o conteúdo de imagem tridimensional tendo um valor máximo de profundidade, exibir o conteúdo de imagem tridimensional e exibir o texto no valor máximo de profundidade.

De acordo ainda com outro aspecto da presente revelação, um sistema para combinar texto com conteúdo de imagem tridimensional inclui meio para receber conteúdo de imagem tridimensional, meio para determinar um valor máximo de profundidade para o conteúdo tridimensional, e meio para combinar texto com o conteúdo de imagem tridimensional no valor máximo de profundidade.

De acordo ainda com um aspecto adicional da presente revelação, um sistema para exibir texto com conteúdo de imagem tridimensional inclui meio para receber conteúdo de imagem tridimensional e texto, o conteúdo de imagem tridimensional tendo um valor máximo de profundidade, meio para exibir o conteúdo de imagem tridimensional e meio para exibir o texto no valor máximo de profundidade.

Breve descrição dos desenhos

Esses e outros aspectos, características e vantagens da presente revelação serão descritos ou se tornarão evidentes a partir da seguinte descrição detalhada das modalidades preferidas, que deve ser lida com relação aos desenhos em anexo.

Nos desenhos, onde numerais de referência similares indicam elementos similares em todas as vistas:

A figura 1 é uma ilustração exemplar de um sistema para combinar texto com conteúdo tridimensional de acordo com um aspecto da presente revelação;

A figura 2 ilustra um exemplo de uma imagem 2D e um mapa de profundidade as-

sociado à imagem 2D;

A figura 3 ilustra um exemplo de texto adicionado à imagem 2D e o mapa de profundidade associado à imagem 2D de acordo com a presente revelação;

5 A figura 4 é um fluxograma que ilustra um processo de inserção de legenda para ouvintes off-line de acordo com a presente revelação;

A figura 5 é um fluxograma que ilustra um processo de inserção de legenda para ouvintes on-line de acordo com a presente revelação;

A figura 6 ilustra um processo de inserção e detecção de legenda para ouvintes on-line de acordo com a presente revelação; e

10 A figura 7 ilustra um exemplo de texto combinado com um par estéreo de acordo com a presente revelação.

Deve ser entendido que o(s) desenho(s) é (são) para fins de ilustrar os conceitos da revelação e não é (são) necessariamente a única configuração possível para ilustrar a revelação.

15 Descrição detalhada de modalidades preferidas

Deve ser entendido que os elementos mostrados nas figuras podem ser implementados em várias formas de hardware, software ou combinações dos mesmos. Preferivelmente, esses elementos são implementados em uma combinação de hardware e software em um ou mais dispositivos de propósito geral apropriadamente programados, que podem incluir um processador, memória e interfaces de entrada/saída.

20 A presente descrição ilustra os princípios da presente revelação. Será desse modo reconhecido que aqueles versados na técnica serão capazes de idealizar vários arranjos que, embora não explicitamente descritos ou mostrados aqui, incorporam os princípios da revelação e são incluídos em seu espírito e escopo.

25 Todos os exemplos e linguagem condicional mencionados aqui são para fins pedagógicos para auxiliar o leitor a entender os princípios da revelação e os conceitos contribuídos pelo inventor para incrementar a técnica, e devem ser interpretados como sendo sem limitação a tais exemplos e condições especificamente mencionados.

30 Além disso, todas as afirmações aqui que mencionam princípios, aspectos e modalidades da revelação, bem como exemplos específicos das mesmas, pretendem abranger equivalentes tanto estruturais como funcionais das mesmas. Adicionalmente, pretende-se que tais equivalentes incluam tanto equivalentes atualmente conhecidos como equivalente desenvolvidos no futuro, isto é, quaisquer elementos desenvolvidos que realizem a mesma função, independente de estrutura.

35 Desse modo, por exemplo, será reconhecido por aqueles versados na técnica que os diagramas de blocos apresentados aqui representam vistas conceptuais de conjuntos de circuitos ilustrativos que incorporam os princípios da revelação. Similarmente, será reconhe-

cido que quaisquer fluxogramas, diagramas de fluxo, diagramas de transição de estado, pseudocódigo, e similar representam vários processos que podem ser substancialmente representados em meios legíveis por computador e desse modo executados por um computador ou processador, quer ou não esse computador ou processador seja explicitamente

5 mostrado.

As funções dos vários elementos mostrados nas figuras podem ser fornecidas através do uso de hardware dedicado bem como hardware capaz de executar software em associação a software apropriado. Quando fornecido por um processador, as funções podem ser fornecidas por um processador dedicado único, por um processador compartilhado único, ou por uma pluralidade de processadores individuais, alguns dos quais podem ser compartilhados. Além disso, o uso explícito do termo “processador” ou “controlador” não deve

10 ser interpretado como se referindo exclusivamente a hardware capaz de executar software, e pode incluir implicitamente, sem limitação, hardware de processador de sinais digitais “DSP”, memória somente de leitura “ROM” para armazenar software, memória de acesso

15 aleatório “RAM” e armazenagem não volátil.

O outro hardware, convencional e/ou customizado, também pode ser incluído. Similarmente, quaisquer comutações mostradas nas figuras são somente conceptuais. Sua função pode ser realizada através da operação de lógica de programa, através de lógica dedicada, através da interação de controle de programa e lógica dedicada, ou mesmo manualmente, a técnica específica sendo selecionável pelo implementador como entendido mais

20 especificamente a partir do contexto.

Nas reivindicações do presente, qualquer elemento expresso como meio para executar uma função específica pretende abranger qualquer modo de executar aquela função incluindo, por exemplo, a) uma combinação de elementos de circuito que executa aquela

25 função ou b) software em qualquer forma, incluindo, portanto, firmware, microcódigo ou similar, combinado com conjunto de circuitos apropriado para executar aquele software para realizar a função. A revelação como definido por tais reivindicações reside no fato de que as funcionalidades fornecidas pelos vários meios mencionados são combinadas e unidas no modo que as reivindicações exigem. Desse modo, é considerado que qualquer meio que

30 possa fornecer essas funcionalidades é equivalente àqueles mostrados aqui.

Com referência agora à figura 1, componentes de sistema exemplares 10, de acordo com uma modalidade da presente revelação, são mostrados. Um dispositivo de varredura 12 pode ser fornecido para varrer cópias de filme 14, por exemplo, negativos de filme original de câmera, em um formato digital, por exemplo, um formato Cineon ou arquivos de

35 Society of Motion Picture and Television Engineers (SMPTE) Digital Picture Exchange (DPX). O dispositivo de varredura 12 pode compreender, por exemplo, um telecine ou qualquer dispositivo que gerará uma saída de vídeo a partir do filme como, por exemplo, um Arri

LocPro™ com saída de vídeo. Alternativamente, arquivos a partir do processo pós-produção ou cinema digital 16 (por exemplo, arquivos já em forma legível por computador) podem ser utilizados diretamente. Fontes em potencial de arquivos legíveis por computador são editores AVID™, arquivos DPX, fitas D5, etc. Além disso, o conteúdo 3D (por exemplo, conteúdo estereoscópico ou imagens 2D e mapas de profundidade associados) pode ser fornecido por um dispositivo de captura 18 e arquivos de texto 20 (por exemplo, arquivos de legenda para ouvintes ou legenda para deficientes auditivos) podem ser criados a partir de um script e fornecidos ao sistema pelo supervisor de legenda para ouvintes.

As cópias de filme varridas, imagens de filme digital e/ou conteúdo 3D bem como os arquivos de texto podem ser inseridos em um dispositivo pós-processamento 22, por exemplo, um computador. O computador 22 pode ser implementado em qualquer uma das várias plataformas de computador conhecidas tendo hardware como uma ou mais unidades de processamento central (CPU), memória 24 como memória de acesso aleatório (RAM) e/ou memória somente de leitura (ROM) e interface(s) de usuário de entrada/saída (I/O) 26 como um teclado, dispositivo de controle de cursor (por exemplo, um mouse ou manche) e dispositivo de exibição. A plataforma de computador também inclui um sistema operacional e código de instrução micro. Os vários processos e funções descritas aqui podem fazer parte do código de instrução micro ou parte de um programa de aplicação de software (ou uma combinação dos mesmos) que é executado através do sistema operacional. Além disso, vários outros dispositivos periféricos podem ser conectados à plataforma de computador por várias interfaces e estruturas de barramento, como porta paralela, porta serial ou barramento serial universal (USB). Outros dispositivos periféricos podem incluir dispositivos de armazenagem adicionais 28 e uma impressora 30. A impressora 30 pode ser empregada para imprimir uma versão revisada do filme 32, por exemplo, uma versão estereoscópica do filme, onde texto foi inserido em uma cena ou uma pluralidade de cenas utilizando as técnicas de inserção de texto descritas abaixo. Adicionalmente, um arquivo digital 34 do vídeo ou filme revisado pode ser gerado e fornecido a um dispositivo de exibição 3D de modo que o conteúdo 3D e texto inserido possam ser vistos por um telespectador. Alternativamente, o arquivo digital 34 pode ser armazenado no dispositivo de armazenagem 28.

Um programa de software inclui um módulo de processamento de texto 38 armazenado na memória 24 para combinar texto com conteúdo 3D de acordo com a presente revelação, como discutido em detalhes adicionais abaixo.

Há diversas técnicas para apresentar conteúdo 3D. A mais comum é meio de exibição estereoscópico, que requer vidros ativo ou passivo. Meios de exibição auto-estereoscópicos, utilizando, por exemplo, Lenticular, não requerem vidros e estão se tornando mais disponíveis para entretenimento tanto em casa como profissional. Muitos desses meios de exibição operam no formato 2D + profundidade. Nesse formato, o vídeo 2D e

as informações de profundidade são combinados para criar o efeito 3D.

A presente revelação é dirigida a um método para inserir legendas para ouvintes no vídeo 3D para meios de exibição do tipo estéreo e 2D+profundidade. Para meios de exibição 2D+profundidade, o método proposto insere texto de legenda para ouvintes no mesmo nível que o valor de profundidade mais elevado na imagem. Mais especificamente, o valor de profundidade da legenda para ouvintes inserida pode ser ajustado continuamente para casar com o valor de profundidade maior do mapa de profundidade. Para conteúdo estéreo, o método proposto ajusta o valor de disparidade da legenda para ouvintes na imagem direita. Isso produz legendas para ouvintes mais visualmente agradáveis que não obstruem os efeitos 3D do vídeo.

Legendas para ouvintes podem ser colocadas em um sinal de vídeo em um de dois modos: on-line (ao vivo) ou off-line (pós-produção). Legenda para ouvintes on-line é feita à medida que um evento ocorre. Os exemplos de legenda para ouvintes on-line são programas de notícias de televisão, seminários ao vivo e eventos esportivos. Legendas para ouvintes on-line podem ser feitas a partir de um script, ou na realidade criados em tempo real. Legenda para ouvintes off-line é feita “após o fato” em um estúdio. Os exemplos de legendagem off-line incluem shows de jogos de televisão, videoteipes ou DVDs de filmes, videoteipes de corporações (por exemplo, vídeos de treinamento), filmes fornecidos através de cabo, satélite ou Internet, ou similar. O texto da legenda para ouvintes é criado em um computador, e sincronizado com o vídeo utilizando códigos de tempo. O texto e o vídeo são então transferidos para o videoteipe antes do mesmo ser transmitido ou distribuído.

Na presente revelação, a criação e distribuição de legendas para ouvintes segue, preferivelmente, processos convencionais como conhecidos por aqueles versados na técnica. Por exemplo, um processo convencional é criar um arquivo de texto a partir de um script. O arquivo de texto contém três valores (quadro de início, quadro final, e texto). O texto é então repetido em todos os quadros a partir do quadro de início até o quadro final. A presente revelação é dirigida ao ajuste do valor de profundidade do local de texto de tal modo que o valor de profundidade do local de texto case com o valor de profundidade maior no quadro de vídeo.

Há diversos formatos de conteúdo e meios de exibição no mercado incluindo estereoscópico, holográfico, e auto-estereoscópico entre outros. Com referência agora à figura 2, uma modalidade da presente revelação é dirigida a uma abordagem para inserção de legendas para ouvintes em meios de exibição auto-estereoscópico que operam no formato 2D+profundidade. A figura 2 ilustra um exemplo de formato de conteúdo 2D+profundidade. Mais especificamente, a figura 2 ilustra dois tipos de conteúdos: uma imagem 2D 40 e um mapa de profundidade 42 da imagem 2D. O mapa de profundidade 42 define o valor de profundidade em cada pixel na imagem 2D 40 com pixels claros que representam pontos pró-

ximos ao telespectador, e pixels escuros que representam pontos distantes do telespectador.

Como discutido acima, há dois modos para inserir legendas para ouvintes: inserção on-line para conteúdo ao vivo e inserção off-line para conteúdo pós-produção. Como discutido abaixo, os métodos propostos da presente revelação são dirigidos à inserção de legenda para ouvintes tanto off-line como on-line.

Com referência agora à figura 3, um exemplo de uma caixa de texto 50 inserida em um mapa de profundidade 46 e o texto 48 adicionado à imagem 2D 40 é mostrado. A caixa de texto 48 é o texto de legenda para ouvintes, como definido pelo script, por exemplo, enquanto a caixa de texto 50 representa um valor de profundidade constante em cada ponto da caixa de texto.

Com referência agora à figura 4, um processo de inserção off-line 52 da presente revelação é mostrado. Para inserção de legendas para ouvintes off-line, imagens de texto de legenda para ouvintes são criadas e sincronizadas com vídeo 2D utilizando códigos de tempo em produção posterior. Os valores de profundidade do texto inserido são determinados por varredura, na etapa 54, do vídeo 3D e cálculo do valor max. da profundidade para cada quadro durante a criação de conteúdo. Uma nova caixa de texto é então inserida, na etapa 56, no local de legenda para ouvintes com valor de profundidade igual ao valor max. de profundidade do quadro, e na etapa 58, a legenda para ouvintes é adicionada à imagem 2D 44. Esse processo deve ser feito para a duração do intervalo de tempo definido para a legenda para ouvintes. Deve ser observado que as etapas 56 e 58 podem ser realizadas em qualquer ordem e podem ser executadas preferivelmente simultaneamente.

Com referência agora à figura 5, é mostrado um fluxograma da presente revelação que ilustra um processo de inserção on-line 60. No processamento on-line, o local das legendas para ouvintes não é sabido antecipadamente e conseqüentemente o valor de profundidade das legendas para ouvintes não pode ser determinado do mesmo modo como descrito para processamento off-line 52. Assim que o texto de legenda para ouvintes é inserido, na etapa 62, o mapa de profundidade do quadro de início de legenda para ouvintes é varrido para determinar o valor max. de profundidade e na etapa 64, o texto de legenda para ouvintes é inserido no valor max. de profundidade e, na etapa 66, a legenda para ouvintes é adicionada à imagem 2D. Deve ser observado que as etapas 64 e 66 podem ser realizadas em qualquer ordem e podem ser preferivelmente realizadas simultaneamente. Posteriormente, na etapa 68, uma determinação é feita com relação a se existem recursos adicionais de processamento. Dependendo do processamento disponível, a legenda para ouvintes pode ser fixa, na etapa 70, no valor de profundidade do primeiro quadro quando processamento adicional não está disponível ou os valores de profundidade dos quadros seguintes podem ser determinados repetindo as etapas de processamento on-line 62-66 quando proces-

samento adicional está disponível.

Com referência agora à figura 6, é mostrado um fluxograma da presente revelação que ilustra o processamento 72 de imagens 2D tendo legendas para ouvintes inseridas. Há casos onde legendas para ouvintes já estão inseridas na imagem 2D como se o conteúdo 3D fosse convertido do conteúdo 2D. Para esses casos, o local de legendas para ouvintes pode ser identificado, na etapa 74, por detectores de região de legenda para ouvintes, que são capazes de detectar e localizar as regiões de legenda para ouvintes em um quadro utilizando informações de cor e textura. Detecção de região de legenda para ouvintes tem sido uma direção de pesquisa ativa na pesquisa de processamento de vídeo. De acordo com a literatura atual, para alguns vídeos, como vídeos de notícias, detectores de região de legenda para ouvintes podem obter precisão de localização acima de 95%. Portanto, detectores de região de legenda para ouvintes devem ser seguros o bastante para inserção de legenda para ouvintes 3D. Após localização da área de legenda para ouvintes (isto é, a coordenada da caixa de texto é determinada), na etapa 74, e o texto de legenda para ouvintes é isolado (isto é, os pixels específicos da legenda para ouvintes são determinados), na etapa 76, a partir da imagem, o mapa de profundidade do quadro de início de legenda para ouvintes é buscado (por exemplo, varrido) para determinar, na etapa 78, o valor max. de profundidade. A seguir, na etapa 80, o texto de legenda para ouvintes é inserido no valor max. de profundidade. Posteriormente, as etapas de processo de inserção on-line 66-70 mostradas na figura 5, podem ser aplicadas.

Com referência agora à figura 7, a presente revelação também pode ser estendida para cobrir conteúdo estereoscópico 82. Para conteúdo estereoscópico o texto na imagem de olho esquerdo ou direito é deslocado para casar com o valor de profundidade maior na imagem estereoscópica. Por exemplo, o texto 88 pode ser fixo na imagem de olho esquerdo 84 porém ajustado ou variado na imagem de olho direito 86. A variação do texto 90 na imagem de olho direito 86 é proporcional à disparidade do par estéreo. O valor de disparidade é inversamente proporcional ao valor de profundidade.

A variação no olho é um deslocamento na direção horizontal. Um deslocamento negativo (fora do texto da tela) é preferível para a maioria das aplicações. Entretanto a presente revelação permite deslocamentos tanto negativo como positivo do texto. O valor de deslocamento mínimo permitido é igual ao valor positivo máximo visualmente aceitável e o valor de deslocamento máximo permitido é igual ao valor negativo máximo visualmente aceitável. A figura 7 mostra um exemplo de par estéreo com um valor de deslocamento de 10 pixels para o texto 90 na imagem de olho direito 86.

Deve ser observado que, de acordo com a presente revelação, é desejável combinar texto com conteúdo 3D (por exemplo, conteúdo estereoscópico ou imagens 2D e mapas de profundidade associados) de tal modo que o texto seja ocasional ou continuamente posi-

cionado no valor máximo de profundidade do conteúdo 3D. Abaixo, várias abordagens para adquirir informações de profundidade a partir do conteúdo 3D são discutidas adicionalmente.

A aquisição de informações de profundidade pode ser feita utilizando técnicas ativa ou passiva. Abordagens passivas adquirem geometria 3D a partir de imagens ou vídeos feitos sob condições de iluminação regular. A geometria 3D é computada utilizando as características geométricas ou fotométricas extraídas de imagens e vídeos. Abordagens ativas utilizam fontes de luz especial, como laser, luz de estrutura ou luz infravermelha. Computam a geometria com base na reposta dos objetos e cenas à luz especial projetada sobre a superfície.

Abordagens de vista única recuperam geometria 3D utilizando uma imagem tirada de um ponto de vista de câmera única. Os exemplos incluem profundidade e estéreo fotométrico a partir de desfocagem. Abordagens de múltiplas vistas recuperam geometria 3D a partir de múltiplas imagens tiradas de pontos de vista de câmeras múltiplas, resultadas de movimento de objeto, ou com diferentes posições de fonte de luz. O casamento de estéreo é um exemplo de recuperação 3D de múltiplas vistas por casamento dos pixels na imagem esquerda e imagem direita no par de estéreo para obter as informações de profundidade dos pixels.

Os métodos geométricos recuperam geometria 3D por detectar características geométricas como cantos, linhas ou contornos em imagens única ou múltiplas. A relação espacial entre os cantos, linhas ou contornos extraídos pode ser utilizada para inferir as coordenadas 3D dos pixels em imagens. Os métodos fotométricos recuperam geometria 3D com base no sombreamento ou sombra dos patches de imagem resultados da orientação da superfície de cena.

Para a aplicação da presente revelação, há três tipos possíveis de conteúdo: conteúdo gerado por computador, conteúdo estéreo e conteúdo 2D. Para conteúdo gerado por computador, como utilizado em animação, informações de profundidade são disponíveis com processamento muito limitado. Para conteúdo de estéreo, a imagem direita e esquerda pode ser utilizada para gerar a profundidade por casar o pixel na imagem esquerda com aquele na imagem direita. O caso mais complexo é aquele de conteúdo 2D. A maioria das técnicas atuais envolve processamento manual extenso e conseqüentemente devem ser feitas off-line. Para aplicações de cinema digital, o conteúdo 2D é convertido em par estéreo para reprodução em cinemas digitais. Após aquisição do par estéreo, técnicas de estéreo podem ser utilizadas para obter um mapa de profundidade. Em geral para aplicações de legenda para ouvintes mapas de profundidade altamente precisos e densos não são normalmente necessários.

Embora as modalidades que incorporam os ensinamentos da presente revelação tenham sido mostradas e descritas em detalhe aqui, aqueles versados na técnica podem

- facilmente idealizar muitas outras modalidades variadas que ainda incorporam esses ensinamentos. Tendo descrito modalidades preferidas para um sistema e método para processamento de imagem paralela em um ambiente de computação ligado em rede com esquemas de divisão de dados de imagem ótimos (que pretendem ser ilustrativos e não limitadores). Observa-se que modificações e variações podem ser feitas por pessoas versadas na técnica à luz dos ensinamentos acima. Portanto, deve ser entendido que alterações podem ser feitas nas modalidades específicas da revelação revelada que estão compreendidas no escopo da revelação como delineado pelas reivindicações apensas.
- 5

REIVINDICAÇÕES

1. Método para combinar texto com conteúdo de imagem tridimensional, o método sendo **CARACTERIZADO** pelo fato de que compreende as etapas de:

receber (54) conteúdo de imagem tridimensional;

5 determinar (54) um valor máximo de profundidade para o conteúdo tridimensional; e
 combinar (58) texto com o conteúdo de imagem tridimensional no valor máximo de profundidade.

2. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que a etapa de receber conteúdo de imagem tridimensional inclui receber (54) uma imagem bidimensional (40) e um mapa de profundidade (42).
10

3. Método, de acordo com a reivindicação 2, **CARACTERIZADO** pelo fato de que a etapa de determinar (54) um valor máximo de profundidade inclui detectar qual objeto no mapa de profundidade tem o valor máximo de profundidade.

4. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que a etapa de combinar (58) texto com o conteúdo tridimensional inclui sobrepor o texto na imagem bidimensional e posicionar o texto no mapa de profundidade no valor máximo de profundidade.
15

5. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o conteúdo de imagem tridimensional inclui uma pluralidade de quadros e as etapas de determinar (62) o valor máximo de profundidade e combinar (64, 66) o texto com o conteúdo de imagem tridimensional no valor máximo de profundidade ocorrem para cada quadro.
20

6. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o conteúdo de imagem tridimensional inclui uma pluralidade de quadros e as etapas de determinar (62) o valor máximo de profundidade e combinar (64, 66) o texto com o conteúdo de imagem tridimensional no valor máximo de profundidade ocorrem para um número menor do que todos da pluralidade de quadros.
25

7. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que compreende ainda as etapas de:

determinar (74) se o conteúdo tridimensional contém texto;

30 isolar (76) o texto a partir do conteúdo tridimensional; e

 combinar (78, 80) o texto isolado com o conteúdo tridimensional no valor máximo de profundidade.

8. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que o texto é um entre legendas para ouvintes, legendagem oculta e legendagem aberta.

9. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que a etapa de determinar o valor máximo de profundidade para o conteúdo tridimensional inclui detectar o valor máximo de profundidade de um objeto em uma imagem estereoscópica
35

(82), a imagem estereoscópica (82) incluindo uma imagem de olho esquerdo (84) e uma imagem de olho direito (86).

10. Método, de acordo com a reivindicação 9, **CARACTERIZADO** pelo fato de que a etapa de combinar texto com a imagem tridimensional inclui:

5 sobrepor o texto (88) na imagem de olho esquerdo (84);
 sobrepor o texto (90) na imagem de olho direito (86); e
 deslocar o texto (90) na imagem de olho direito (86) de tal modo que o texto de olho direito e olho esquerdo combinado é exibível no valor máximo de profundidade da imagem estereoscópica.

10 11. Método de exibir texto com conteúdo de imagem tridimensional, o método sendo **CARACTERIZADO** pelo fato de que compreende as etapas de:

 receber (18, 20) conteúdo de imagem tridimensional e texto, o conteúdo de imagem tridimensional tendo um valor máximo de profundidade;

 exibir (36) o conteúdo de imagem tridimensional; e

15 exibir (36) o texto no valor máximo de profundidade.

12. Método, de acordo com a reivindicação 11, **CARACTERIZADO** pelo fato de que compreende ainda a etapa de:

 determinar (54) o valor máximo de profundidade do conteúdo de imagem tridimensional.

20 13. Método, de acordo com a reivindicação 12, **CARACTERIZADO** pelo fato de que a etapa de determinar (54) compreende detectar qual objeto no conteúdo de imagem tridimensional tem o valor máximo de profundidade.

14. Método, de acordo com a reivindicação 12, **CARACTERIZADO** pelo fato de que o conteúdo de imagem tridimensional inclui uma pluralidade de quadros e as etapas de determinar (62) o valor máximo de profundidade e exibir (36) o texto no valor máximo de profundidade ocorrem para cada quadro.

25 15. Método, de acordo com a reivindicação 12, **CARACTERIZADO** pelo fato de que o conteúdo de imagem tridimensional inclui uma pluralidade de quadros e as etapas de determinar (62) o valor máximo de profundidade e exibir (36) o texto no valor máximo de profundidade ocorrem para um número menor do que todos da pluralidade de quadros.

30 16. Método, de acordo com a reivindicação 11, **CARACTERIZADO** pelo fato de que o texto é um entre legendas para ouvintes, legendagem oculta e legendagem aberta.

17. Método, de acordo com a reivindicação 11, **CARACTERIZADO** pelo fato de que compreende ainda as etapas de:

35 determinar (74) se o conteúdo tridimensional contém texto;

 isolar (76) o texto a partir do conteúdo tridimensional; e

 exibir (36) o texto isolado no valor máximo de profundidade.

18. Método, de acordo com a reivindicação 11, **CARACTERIZADO** pelo fato de que a etapa de determinar o valor máximo de profundidade para o conteúdo tridimensional inclui detectar o valor máximo de profundidade de um objeto em uma imagem estereoscópica (82), a imagem estereoscópica incluindo uma imagem de olho esquerdo (84) e uma imagem de olho direito (86).

19. Método, de acordo com a reivindicação 18, **CARACTERIZADO** pelo fato de que a etapa de combinar texto com a imagem tridimensional inclui:

sobrepor texto (88) na imagem de olho esquerdo (84);

sobrepor texto (90) na imagem de olho direito (86); e

deslocar o texto (90) na imagem de olho direito (86) de tal modo que o texto de olho direito e olho esquerdo combinado é exibível no valor máximo de profundidade da imagem estereoscópica.

20. Sistema para combinar texto com conteúdo de imagem tridimensional, o sistema sendo **CARACTERIZADO** pelo fato de que compreende:

meio para receber (54) conteúdo de imagem tridimensional;

meio para determinar (54) um valor máximo de profundidade para o conteúdo tridimensional; e

meio para combinar (58) texto com o conteúdo de imagem tridimensional no valor máximo de profundidade.

21. Sistema, de acordo com a reivindicação 20, **CARACTERIZADO** pelo fato de que o meio para receber conteúdo de imagem tridimensional inclui meio para receber (54) uma imagem bidimensional (40) e um mapa de profundidade (42).

22. Sistema, de acordo com a reivindicação 21, **CARACTERIZADO** pelo fato de que o meio para determinar (54) um valor máximo de profundidade inclui meio para detectar qual objeto no mapa de profundidade tem o valor máximo de profundidade.

23. Sistema, de acordo com a reivindicação 20, **CARACTERIZADO** pelo fato de que o meio de combinar (58) texto com o conteúdo tridimensional inclui meio para sobrepor o texto na imagem bidimensional e meio para posicionar o texto no mapa de profundidade no valor máximo de profundidade.

24. Sistema, de acordo com a reivindicação 20, **CARACTERIZADO** pelo fato de que o conteúdo de imagem tridimensional inclui uma pluralidade de quadros e o meio para determinar (62) o valor máximo de profundidade e meio para combinar (64, 66) o texto com o conteúdo de imagem tridimensional no valor máximo de profundidade ocorrem para cada quadro.

25. Sistema, de acordo com a reivindicação 20, **CARACTERIZADO** pelo fato de que o conteúdo de imagem tridimensional inclui uma pluralidade de quadros e o meio para determinar (62) o valor máximo de profundidade e meios para combinar (64, 66) o texto com

o conteúdo de imagem tridimensional no valor máximo de profundidade ocorrem para um número menor do que todos da pluralidade de quadros.

26. Sistema, de acordo com a reivindicação 20, **CARACTERIZADO** pelo fato de que compreende ainda:

- 5 meio para determinar (74) se o conteúdo tridimensional contém texto;
- meio para isolar (76) o texto a partir do conteúdo tridimensional; e
- meios para combinar (78, 80) o texto isolado com o conteúdo tridimensional no valor máximo de profundidade.

27. Sistema, de acordo com a reivindicação 20, **CARACTERIZADO** pelo fato de
10 que o texto é um entre legendas para ouvintes, legendagem oculta e legendagem aberta.

28. Sistema, de acordo com a reivindicação 20, **CARACTERIZADO** pelo fato de que o meio para determinar o valor máximo de profundidade para o conteúdo tridimensional inclui meio para detectar o valor máximo de profundidade de um objeto em uma imagem estereoscópica (82), a imagem estereoscópica (82) incluindo uma imagem de olho esquerdo
15 (84) e uma imagem de olho direito (86).

29. Sistema, de acordo com a reivindicação 28, **CARACTERIZADO** pelo fato de que a etapa de combinar texto com a imagem tridimensional inclui:

- meio para sobrepor o texto (88) na imagem de olho esquerdo (84);
- meio para sobrepor o texto (90) na imagem de olho direito (86); e
- 20 meio para deslocar o texto (90) na imagem de olho direito (86) de tal modo que o texto de olho direito e olho esquerdo combinado é exibível no valor máximo de profundidade da imagem estereoscópica.

30. Sistema para exibir texto com conteúdo de imagem tridimensional, o sistema sendo **CARACTERIZADO** pelo fato de que compreende:

- 25 meio para receber (18, 20) conteúdo de imagem tridimensional e texto, o conteúdo de imagem tridimensional tendo um valor máximo de profundidade;
- meio para exibir (36) o conteúdo de imagem tridimensional; e
- meio para exibir (36) o texto no valor máximo de profundidade.

31. Sistema, de acordo com a reivindicação 30, **CARACTERIZADO** pelo fato de
30 que compreende ainda:

 meio para determinar (54) o valor máximo de profundidade do conteúdo de imagem tridimensional.

32. Sistema, de acordo com a reivindicação 31, **CARACTERIZADO** pelo fato de que o meio para determinar (54) compreende meio para detectar qual objeto no conteúdo de
35 imagem tridimensional tem o valor máximo de profundidade.

33. Sistema, de acordo com a reivindicação 31, **CARACTERIZADO** pelo fato de que o conteúdo de imagem tridimensional inclui uma pluralidade de quadros e o meio para

determinar (54) o valor máximo de profundidade e meio para exibir (36) o texto no valor máximo de profundidade ocorrem para cada quadro.

34. Sistema, de acordo com a reivindicação 31, **CARACTERIZADO** pelo fato de que o conteúdo de imagem tridimensional inclui uma pluralidade de quadros e o meio para determinar (54) o valor máximo de profundidade e o meio para exibir (36) o texto no valor máximo de profundidade operam em um número menor do que todos da pluralidade de quadros.

35. Sistema, de acordo com a reivindicação 30, **CARACTERIZADO** pelo fato de que o texto é um entre legendas para ouvintes, legendagem oculta e legendagem aberta.

36. Sistema, de acordo com a reivindicação 30, **CARACTERIZADO** pelo fato de que compreende ainda:

meio para determinar (74) se o conteúdo tridimensional contém texto;

meio para isolar (76) o texto a partir do conteúdo tridimensional; e

meio para exibir (36) o texto isolado no valor máximo de profundidade.

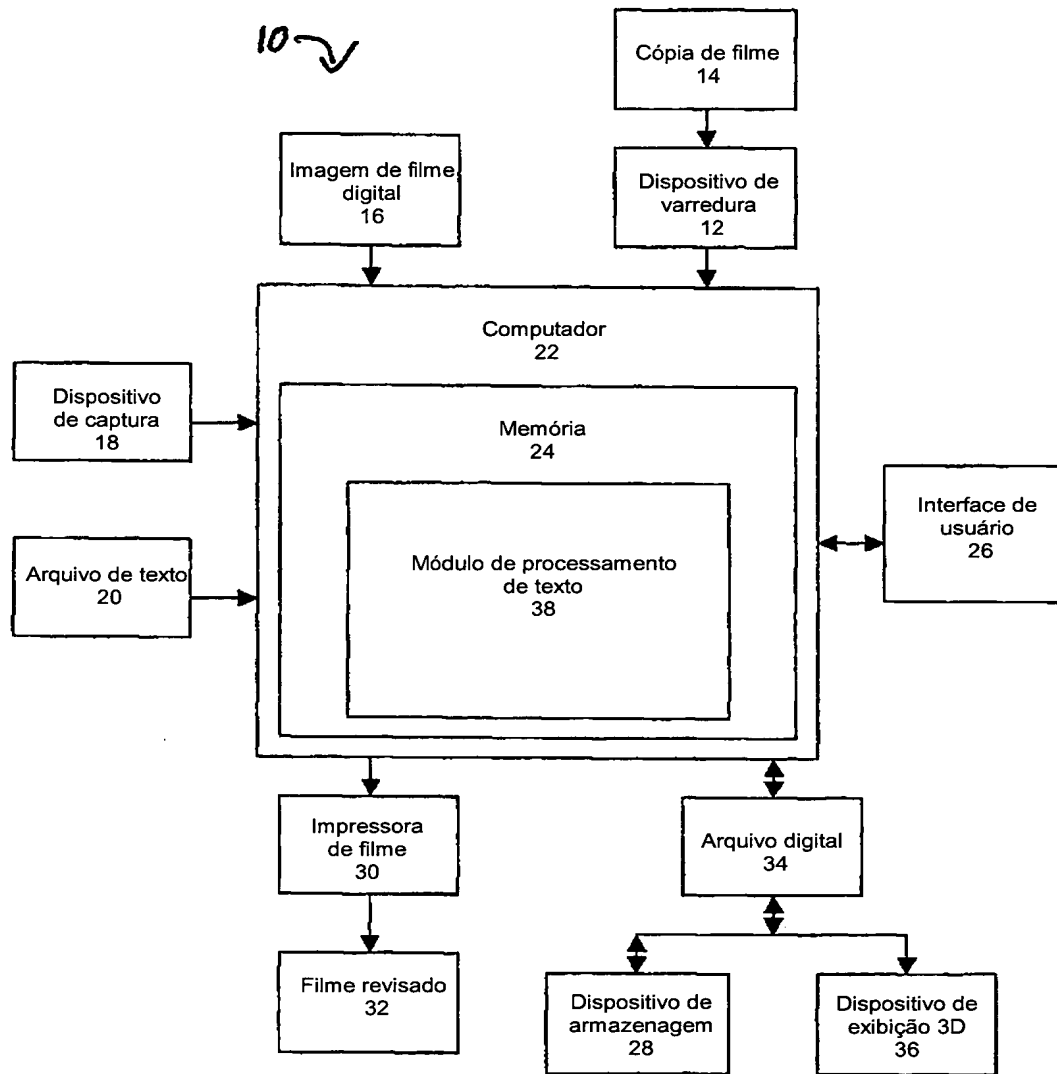
37. Sistema, de acordo com a reivindicação 30, **CARACTERIZADO** pelo fato de que o meio para determinar o valor máximo de profundidade para o conteúdo tridimensional inclui meio para detectar o valor máximo de profundidade de um objeto em uma imagem estereoscópica (82), a imagem estereoscópica incluindo uma imagem de olho esquerdo (84) e uma imagem de olho direito (86).

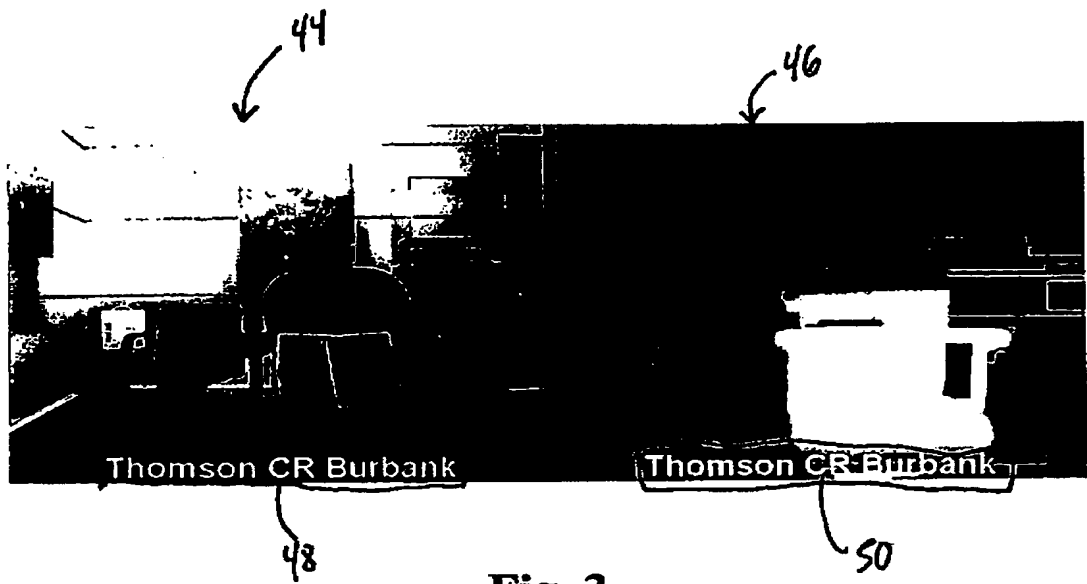
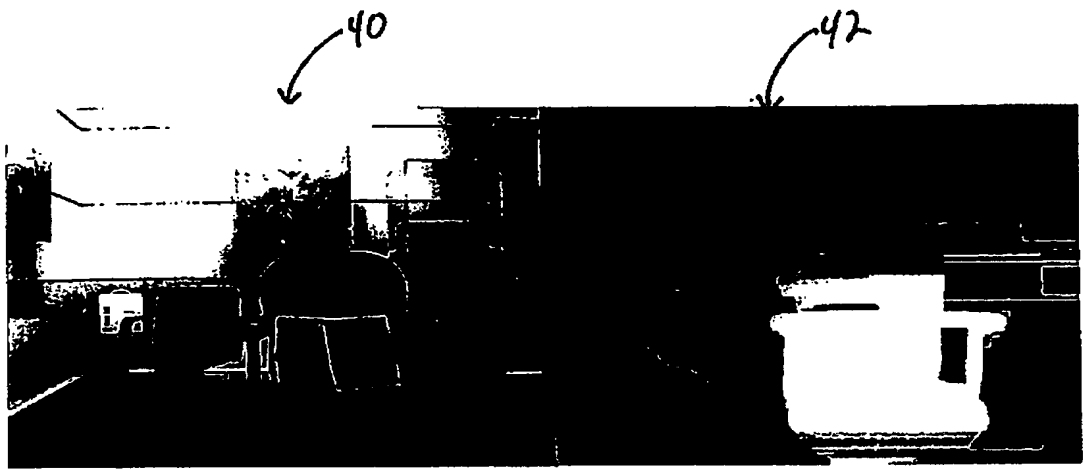
38. Sistema, de acordo com a reivindicação 37, **CARACTERIZADO** pelo fato de que o meio para combinar texto com a imagem tridimensional inclui:

meio para sobrepor texto (88) na imagem de olho esquerdo (84);

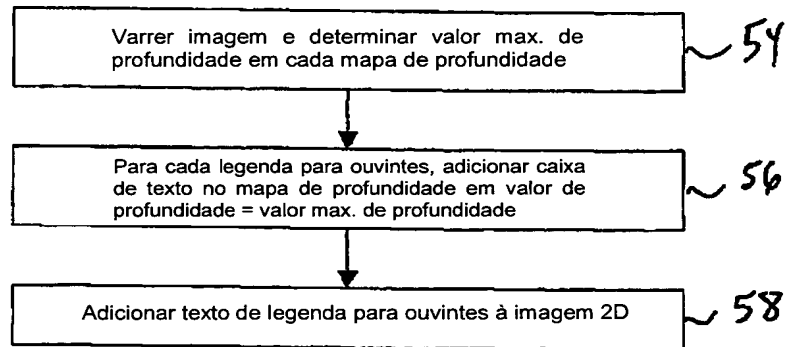
meio para sobrepor texto (90) na imagem de olho direito (86); e

meio para deslocar o texto (90) na imagem de olho direito (86) de tal modo que o texto de olho direito e olho esquerdo combinado é exibível no valor máximo de profundidade da imagem estereoscópica.

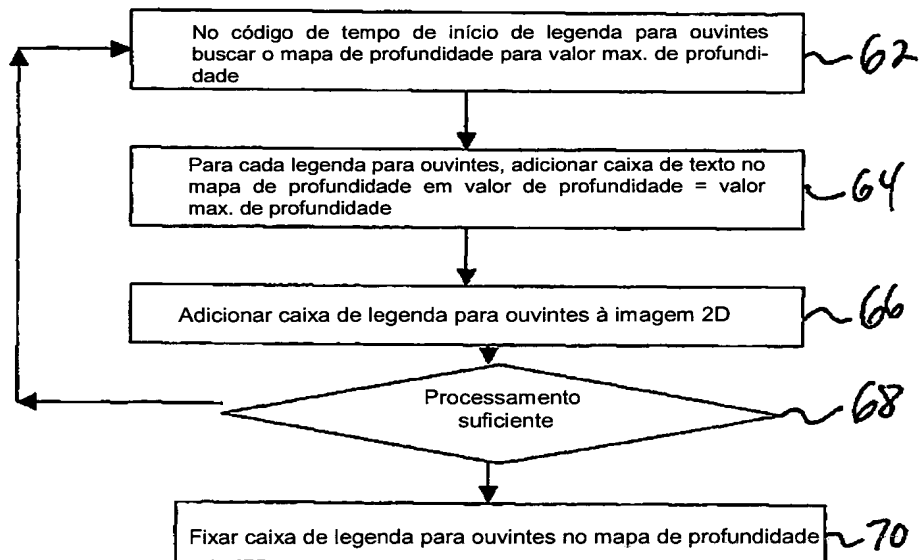
**Fig. 1**

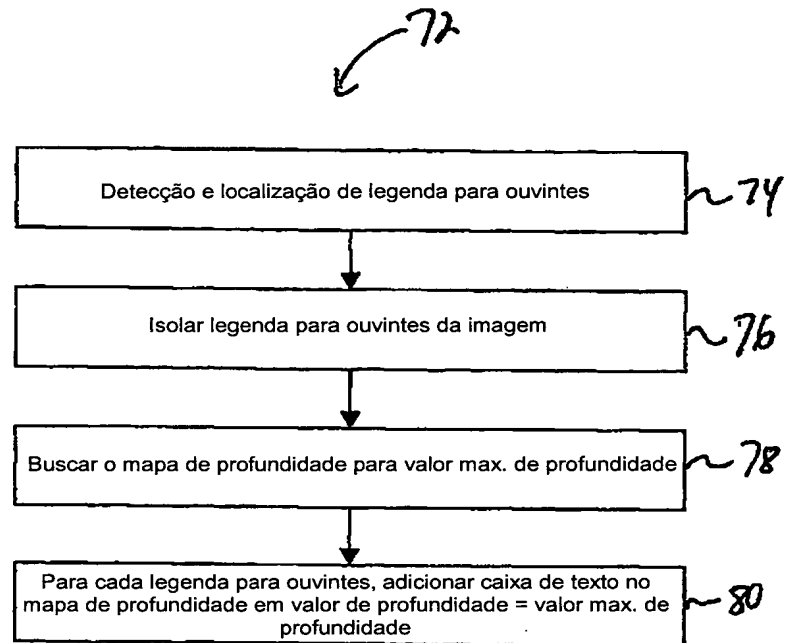
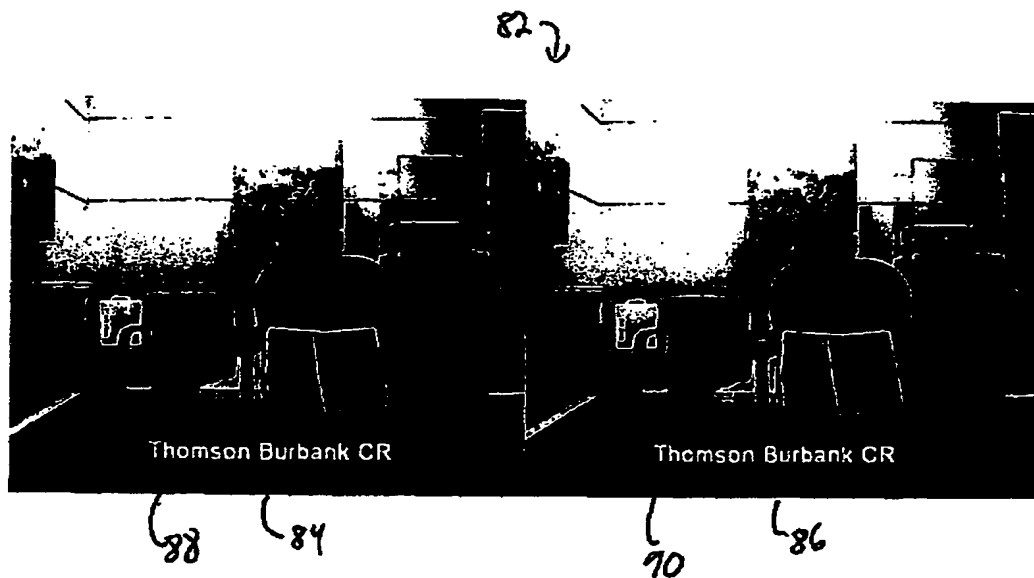
**Fig. 3**

52 ↘

**Fig. 4**

60 ↘

**Fig. 5**

**Fig. 6****Fig. 7**

RESUMO

“SISTEMA E MÉTODO PARA COMBINAR TEXTO COM CONTEÚDO TRIDIMENSIONAL”

Um sistema (10) e método (52, 60, 72) para combinar e/ou exibir texto com conteúdo tridimensional (3D). O sistema (10) e método (52, 60, 72) inserem texto no mesmo nível
5 que o valor de profundidade mais elevado no conteúdo 3D. Um exemplo de conteúdo 3D é uma imagem bidimensional (44) e um mapa de profundidade associado (46). Nesse caso, o valor de profundidade do texto inserido (50) é ajustado para casar com o valor de profundidade maior do mapa de profundidade dado. Outro exemplo de conteúdo 3D é uma pluralidade de imagens bidimensionais e mapas de profundidade associados. Nesse caso, o valor
10 de profundidade do texto inserido é continuamente ajustado para casar com o valor de profundidade maior de um mapa de profundidade dado. Um exemplo adicional de conteúdo 3D é conteúdo estereoscópico (82) tendo uma imagem de olho direito (86) e uma imagem de olho esquerdo (84). Nesse caso o texto (88, 90) em uma das imagens de olho esquerdo (84) e imagem de olho direito (86) são deslocados para casar com o valor de profundidade maior
15 na imagem estereoscópica. Ainda outro exemplo de conteúdo 3D é conteúdo estereoscópico tendo uma pluralidade de imagens de olho direito e imagens de olho esquerdo. Nesse caso o texto em uma das imagens de olho esquerdo ou imagens de olho direito é continuamente deslocado para casar com o valor de profundidade maior nas imagens estereoscópicas. Como resultado, o sistema (10) e método (52, 60, 72) da presente revelação produzem
20 texto combinado com conteúdo 3D onde o texto não obstrui os efeitos 3D no conteúdo 3D e não cria fadiga visual quando visto por um telespectador.