

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 522 589**

51 Int. Cl.:

G08B 23/00 (2006.01)

G08B 31/00 (2006.01)

G08B 13/196 (2006.01)

G06K 9/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **08.02.2008 E 08729422 (9)**

97 Fecha y número de publicación de la concesión europea: **30.07.2014 EP 2118864**

54 Título: **Sistema de reconocimiento conductual**

30 Prioridad:

08.02.2007 US 888777 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

17.11.2014

73 Titular/es:

**BEHAVIORAL RECOGNITION SYSTEMS, INC.
(100.0%)
2100 WEST LOOP S., 9TH FLOOR
HOUSTON, TX 77027, US**

72 Inventor/es:

**EATON, JOHN ERIC;
COBB, WESLEY KENNETH;
URECH, DENNIS GENE;
BLYTHE, BOBBY ERNEST;
FRIEDLANDER, DAVID SAMUEL;
GOTTUMUKKAL, RAJKIRAN KUMAR;
RISINGER, LON WILLIAM;
SAITWAL, KISHOR ADINATH;
SEOW, MING-JUNG;
SOLUM, DAVID MARVIN;
XU, GANG y
YANG, TAO**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 522 589 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Sistema de reconocimiento conductual

Antecedentes de la invención

Campo de la invención

5 La presente invención se refiere, en general, al análisis de vídeo, y más en particular a analizar y aprender una conducta en base a la generación de una corriente de datos de vídeo.

Descripción de la técnica relacionada

10 Algunos sistemas de vídeo vigilancia actualmente disponibles tienen unas capacidades de reconocimiento simples. No obstante, muchos sistemas de vigilancia de este tipo requieren un conocimiento avanzado (antes de que se haya desarrollado un sistema) de las acciones y / o los objetos que los sistemas han de ser capaces de buscar. Debe desarrollarse un código de aplicación subyacente dirigido a conductas “anormales” específicas para que estos sistemas de vigilancia sean operativos y lo bastante funcionales. Dicho de otra forma, a menos que el código subyacente al sistema incluya descripciones de determinadas conductas, el sistema será incapaz de reconocer tales conductas. Además, para conductas distintas, a menudo es necesario desarrollar unos productos de soporte lógico independientes. Esto hace que los sistemas de vigilancia con capacidades de reconocimiento sean laboriosos y prohibitivamente costosos. Por ejemplo, supervisar entradas de aeropuertos en busca de criminales al acecho e identificar nadadores que no se están moviendo en una piscina son dos situaciones distintas y, por lo tanto, pueden requerir el desarrollo de dos productos de soporte lógico distintos que tengan sus respectivas conductas “anormales” previamente codificadas.

20 Los sistemas de vigilancia también pueden diseñarse para memorizar escenas normales y generar una alarma siempre que cambie lo que se considera normal. No obstante, estos tipos de sistemas de vigilancia han de estar pre-programados para saber cuánto de anormal es el cambio. Además, tales sistemas no pueden caracterizar con precisión lo que ha tenido lugar en realidad. En su lugar, estos sistemas determinan que algo previamente considerado “normal” ha cambiado. Por lo tanto, los productos desarrollados de tal forma están configurados para detectar solo un margen limitado de un tipo previamente definido de conducta.

25 El documento WO2006 / 012645 describe un sistema sensitivo que combina detección, rastreo y visualización envolvente de un entorno atestado y abarrotado, tal como un edificio de oficinas, una terminal, u otro sitio cerrado usando una red de cámaras estereoscópicas. Un vigilante supervisa el sitio usando un modelo en 3D en directo, que se actualiza a partir de diferentes direcciones usando los múltiples flujos de vídeo.

30 **Sumario de la invención**

Realizaciones de la presente invención proporcionan un procedimiento y un sistema para analizar y aprender una conducta basándose en un flujo adquirido de tramas de vídeo. Los objetos representados en el flujo se determinan basándose en un análisis de las tramas de vídeo. Cada objeto puede tener un modelo de búsqueda correspondiente, que se usa para rastrear los movimientos de los objetos trama a trama. Se determinan las clases de los objetos y se generan representaciones semánticas de los objetos. Las representaciones semánticas se usan para determinar las conductas de los objetos y para aprender acerca de conductas que tienen lugar en un entorno representado por los flujos de vídeo adquiridos. De esta forma, el sistema aprende con rapidez y en tiempo real conductas normales y anormales para cualquier entorno mediante el análisis de movimientos o actividades o ausencia de las mismas en el entorno e identifica y predice una conducta anormal y sospechosa basándose en lo que se ha aprendido.

40 Una realización particular de la invención incluye un procedimiento para procesar un flujo de tramas de vídeo que registran sucesos dentro de una escena. El procedimiento puede incluir en general recibir una primera trama del flujo. La primera trama incluye datos para una pluralidad de píxeles incluidos en la trama. El procedimiento puede incluir además identificar uno o más grupos de píxeles en la primera trama. Cada grupo representa un objeto dentro de la escena. El procedimiento puede incluir todavía adicionalmente generar un modelo de búsqueda que almacena una o más características asociadas con cada objeto identificado, clasificar cada uno de los objetos usando un clasificador entrenado, rastrear, en una segunda trama, cada uno de los objetos identificados en la primera trama usando el modelo de búsqueda, y suministrar la primera trama, la segunda trama y las clasificaciones de objeto a un motor de aprendizaje automático. El procedimiento puede incluir todavía adicionalmente generar, por el motor de aprendizaje automático, una o más representaciones semánticas de conducta en la que toman parte los objetos en la escena a lo largo de una pluralidad de tramas. El motor de aprendizaje automático puede configurarse en general para aprender patrones de conducta observada en la escena a lo largo de la pluralidad de tramas y para identificar apariciones de los patrones de conducta en la que toman parte los objetos clasificados.

Breve descripción de los dibujos

55 Para la forma en la que las características, ventajas y objetos que se han enunciado en lo que antecede de la

presente invención se obtengan y puedan entenderse con detalle, una descripción más particular de la invención, que se ha resumido brevemente en lo que antecede, puede tenerse por referencia a las realizaciones que se ilustran en los dibujos adjuntos.

5 Ha de observarse, no obstante, que los dibujos adjuntos ilustran solo realizaciones típicas de la presente invención y, por lo tanto, no han de considerarse limitantes de su ámbito, ya que la invención puede admitir otras realizaciones igualmente eficaces.

La figura 1 es un diagrama de bloques de alto nivel de un sistema de reconocimiento de conducta, de acuerdo con una realización de la presente invención.

10 La figura 2 ilustra un diagrama de flujo de un procedimiento para analizar y aprender una conducta basándose en un flujo de tramas de vídeo, de acuerdo con una realización de la presente invención.

La figura 3 ilustra un módulo de segundo plano - primer plano de un motor de visión de ordenador, de acuerdo con una realización de la presente invención.

La figura 4 ilustra un módulo para rastrear objetos de interés en un motor de visión de ordenador, de acuerdo con una realización de la presente invención.

15 La figura 5 ilustra un módulo de estimador / identificador de un motor de visión de ordenador, de acuerdo con una realización de la presente invención.

La figura 6 ilustra un componente de procesador de contexto de un motor de visión de ordenador, de acuerdo con una realización de la presente invención.

20 La figura 7 ilustra un módulo de análisis semántico de un motor de aprendizaje automático, de acuerdo con una realización de la presente invención.

La figura 8 ilustra un módulo de percepción de un motor de aprendizaje automático, de acuerdo con una realización de la presente invención.

25 Las figuras 9A-9C ilustran una secuencia de unas tramas de vídeo en la que un sistema de reconocimiento de conducta detecta una conducta anormal y emite una alerta, de acuerdo con una realización de la presente invención.

Descripción detallada de las realizaciones preferentes

Los sistemas de reconocimiento de conducta de aprendizaje automático, tal como las realizaciones de la invención que se describe en el presente documento, aprenden conductas basándose en información adquirida con el tiempo. En el contexto de la presente invención, se analiza información a partir de un flujo de vídeo (es decir, una secuencia de tramas de vídeo individuales). La presente divulgación describe un sistema de reconocimiento de conducta que aprende a identificar y distinguir entre una conducta normal y anormal dentro de una escena mediante el análisis de movimientos y / o actividades (o ausencia de las mismas) con el tiempo. Las conductas normales / anormales no están previamente definidas o son de codificación fija. En su lugar, el sistema de reconocimiento de conducta que se describe en el presente documento aprende con rapidez lo que es "normal" para cualquier entorno e identifica una conducta anormal y sospechosa basándose en lo que se aprende a través de la supervisión de la ubicación, es decir, mediante el análisis del contenido de vídeo registrado trama a trama.

A continuación, se hace referencia a realizaciones de la invención. No obstante, debería entenderse que la invención no se limita a realización alguna descrita de manera específica. En su lugar, se contempla cualquier combinación de las siguientes características y elementos, ya estén relacionados o no con diferentes realizaciones, para implementar y poner en práctica la invención. Además, en diversas realizaciones la invención proporciona numerosas ventajas frente a la técnica anterior. No obstante, a pesar de que realizaciones de la invención pueden lograr ventajas frente a otras posibles soluciones y / o frente a la técnica anterior, que se consiga o no una ventaja particular por una realización dada no es limitante de la invención. Por lo tanto, los siguientes aspectos, características, realizaciones y ventajas son meramente ilustrativos y no se consideran elementos o limitaciones de las reivindicaciones adjuntas excepto cuando se indique explícitamente en una reivindicación o reivindicaciones. De forma similar, la referencia a "la invención" no deberá interpretarse como una generalización de cualquier materia objeto de la invención que se divulgue en el presente documento y no ha de considerarse que es un elemento o limitación de las reivindicaciones adjuntas excepto cuando se indique explícitamente en una reivindicación o reivindicaciones.

50 Una realización de la invención se implementa como un producto de programa para su uso con un sistema de ordenador. El programa o programas del producto de programa definen funciones de las realizaciones (incluyendo los procedimientos que se describen en el presente documento) y puede estar contenido en una diversidad de medios de almacenamiento legibles por ordenador. Los medios de almacenamiento legibles por ordenador ilustrativos incluyen, pero sin limitarse a: (i) medios de almacenamiento no grabables (por ejemplo, dispositivos de

memoria de solo lectura dentro de un ordenador tal como discos de CD-ROM legibles por una unidad de CD-ROM) en los que se almacena una información de manera permanente; (ii) medios de almacenamiento grabables (por ejemplo, disquetes flexibles dentro de una unidad de disquete o una unidad de disco duro) en los que se almacena una información alterable. Tales medios de almacenamiento legibles por ordenador; cuando portan unas instrucciones legibles por ordenador que dirigen las funciones de la presente invención, son realizaciones de la presente invención. Otros medios incluyen medios de comunicaciones a través de los cuales se transporta información a un ordenador, tal como a través de un ordenador o una red de telefonía, incluyendo redes de comunicaciones inalámbricas. La última realización incluye de manera específica transmitir información a y desde Internet y otras redes. Tales medios de comunicaciones, cuando portan unas instrucciones legibles por ordenador que dirigen las funciones de la presente invención, son realizaciones de la presente invención. En términos generales, en el presente documento puede hacerse referencia a los medios de almacenamiento legibles por ordenador y a los medios de comunicaciones como medios legibles por ordenador.

En general, las rutinas que se ejecutan para implementar las realizaciones de la invención pueden ser parte de un sistema operativo o una aplicación, componente, programa, módulo, objeto, o secuencia de instrucciones, específico. El programa informático de la presente invención por lo general está compuesto por una multitud de instrucciones que se traducirán por el ordenador nativo a un formato legible por máquina y, por lo tanto, serán instrucciones ejecutables. Así mismo, los programas están compuestos por variables y estructuras de datos que o bien residen de forma local en el programa o se encuentran en memoria o en dispositivos de almacenamiento. Además, diversos programas que se describen en el presente documento pueden identificarse basándose en la aplicación para la que estos se implementan en una realización específica de la invención. No obstante, debería apreciarse que todas las nomenclaturas de programa particulares que siguen se usan meramente por conveniencia y, por lo tanto, la invención no debería limitarse a su uso únicamente en aplicación específica alguna identificada y / o implicada por tal nomenclatura.

Realizaciones de la presente invención proporcionan un sistema de reconocimiento de conducta y un procedimiento para analizar, aprender y reconocer conductas. La figura 1 es un diagrama de bloques de alto nivel del sistema de reconocimiento de conducta 100, de acuerdo con una realización de la presente invención. Tal como se muestra, el sistema de reconocimiento de conducta 100 incluye una entrada de vídeo 105, una red 110, un sistema de ordenador 115, y dispositivos de entrada y de salida 145 (por ejemplo, un monitor, un teclado, un ratón, una impresora, y similares).

La red 110 recibe datos de vídeo (por ejemplo, flujo o flujos de vídeo, imágenes de vídeo, o similares) a partir de la entrada de vídeo 105. La entrada de vídeo 105 puede ser una cámara de vídeo, una VCR, un DVR, un DVD, un ordenador, o similares. Por ejemplo, la entrada de vídeo 105 puede ser una cámara de vídeo estacionaria orientada hacia una determinada área (por ejemplo, una estación de metro) y que registra de forma continua el área y los sucesos que tiene lugar en la misma. En general, se hace referencia al área visible para la cámara como la "escena". La entrada de vídeo 105 puede configurarse para registrar la escena como una secuencia de tramas de vídeo individuales a una velocidad de trama especificada (por ejemplo, 24 tramas por segundo), en la que cada trama incluye un número fijo de píxeles (por ejemplo, 320 x 240). Cada píxel de cada trama especifica un valor de color (por ejemplo, un valor de RGB). Además, puede darse formato al flujo de vídeo usando formatos conocidos tales como, por ejemplo, MPEG2, MJPEG, MPEG4, H.263, H.264, y similares. Tal como se analiza con mayor detalle en lo sucesivo, el sistema de reconocimiento de conducta analiza esta información sin procesar para identificar objetos activos en el flujo, clasifica tales elementos, obtiene una diversidad de metadatos con respecto a las acciones e interacciones de tales elementos, y suministra esta información a un motor de aprendizaje automático. A su vez, el motor de aprendizaje automático puede configurarse para evaluar, aprender y recordar con el tiempo. Además, basándose en el "aprendizaje", el motor de aprendizaje automático puede identificar determinadas conductas como anómalas.

La red 110 puede usarse para transmitir los datos de vídeo registrados por la entrada de vídeo 105 al sistema de ordenador 115. En una realización, la red 110 transmite el flujo recibido de tramas de vídeo al sistema de ordenador 115.

A modo de ilustración, el sistema de ordenador 115 incluye una CPU 120, un almacenamiento 125 (por ejemplo, una unidad de disco, una unidad de disco óptico, una unidad de disquete flexible, y similares), y una memoria 130 que contiene un motor de visión de ordenador 135 y un motor de aprendizaje automático 140. El motor de visión de ordenador 135 puede proporcionar una aplicación de soporte lógico configurada para analizar una secuencia de tramas de vídeo proporcionada por la entrada de vídeo 105. Por ejemplo, en una realización, el motor de visión de ordenador 135 puede configurarse para analizar tramas de vídeo para identificar objetivos de interés, rastrear esos objetivos de interés, deducir propiedades acerca de los objetivos de interés, clasificar estos por categorías, y etiquetar los datos observados. En una realización, el motor de visión de ordenador 135 genera una lista de atributos (tales como la textura, el color, y similares) de los objetos de interés clasificados y proporciona la lista al motor de aprendizaje automático 140. Adicionalmente, el motor de visión de ordenador puede suministrar al motor de aprendizaje automático 140 una diversidad de informaciones acerca de cada objeto rastreado dentro de una escena (por ejemplo, datos cinematográficos, datos de profundidad, color, datos, datos de apariencia, etc.).

El motor de aprendizaje automático 140 recibe las tramas de vídeo y los resultados que se generan por el motor de

visión de ordenador 135. El motor de aprendizaje automático 140 analiza los datos recibidos, construye representaciones semánticas de los sucesos representados en las tramas de vídeo, determina patrones, y aprende a partir de estas conductas observadas a identificar sucesos normales y / o anormales. El motor de visión de ordenador 135 y el motor de aprendizaje automático 140 y sus componentes se describen con mayor detalle en lo sucesivo. Los datos que describen si se ha determinado una conducta / suceso normal / anormal, y / o lo que es tal conducta / suceso, pueden proporcionarse a un dispositivo de salida 145 para emitir alertas, por ejemplo, un mensaje de alerta que se presenta en una pantalla de interfaz GUI.

En general, tanto el motor de visión de ordenador 135 como el motor de aprendizaje automático 140 procesan los datos de vídeo recibidos en tiempo real. No obstante, las escalas de tiempo para procesar información por el motor de visión de ordenador 135 y el motor de aprendizaje automático 140 pueden ser diferentes. Por ejemplo, en una realización, el motor de visión de ordenador 135 procesa los datos de vídeo recibidos trama a trama, mientras que el motor de aprendizaje automático procesa los datos recibidos cada N tramas. Dicho de otra forma, mientras que el motor de visión de ordenador 135 analiza cada trama en tiempo real para obtener un conjunto de información acerca de lo que está teniendo lugar dentro de una trama dada, el motor de aprendizaje automático 140 no está limitado por la velocidad de trama en tiempo real de la entrada de vídeo.

Obsérvese, no obstante, que la figura 1 ilustra meramente una posible disposición del sistema de reconocimiento de conducta 100. Por ejemplo, mientras que la entrada de vídeo 105 se muestra conectada con el sistema de ordenador 115 por medio de la red 110, la red 110 no siempre se encuentra presente o es necesaria (por ejemplo, la entrada de vídeo 105 puede estar conectada directamente con el sistema de ordenador 115). Además, en una realización, el motor de visión de ordenador 135 puede implementarse como una parte de un dispositivo de entrada de vídeo (por ejemplo, como un componente de soporte lógico inalterable cableado directamente a una cámara de vídeo). En un caso de este tipo, las salidas de la cámara de vídeo pueden proporcionarse al motor de aprendizaje automático 140 para el análisis.

La figura 2 ilustra un procedimiento 200 para analizar y aprender una conducta a partir de un flujo de tramas de vídeo, de acuerdo con una realización de la presente invención. Tal como se muestra, el procedimiento 200 comienza en la etapa 205. En la etapa 210, se recibe un conjunto de tramas de vídeo a partir de una fuente de entrada de vídeo. En la etapa 215, las tramas de vídeo pueden procesarse para minimizar el ruido de vídeo, una iluminación de escena irregular o inusual, problemas relacionados con el color, y así sucesivamente. Es decir, el contenido de las tramas de vídeo puede potenciarse para mejorar la visibilidad de las imágenes antes del procesamiento por los componentes de un sistema de reconocimiento de conducta (por ejemplo, el motor de visión de ordenador 135 y el motor de aprendizaje automático 140 que se han analizado en lo que antecede).

En la etapa 220, cada trama de vídeo sucesiva se analiza para identificar y / o actualizar una imagen de primer plano y de segundo plano para su uso durante fases subsiguientes del procedimiento 200. En general, la imagen de segundo plano incluye elementos estacionarios de la escena que se está capturando por la entrada de vídeo (por ejemplo, píxeles que representan una plataforma de una estación de metro), mientras que la imagen de primer plano incluye elementos volátiles capturados por la entrada de vídeo (por ejemplo, píxeles que representan un hombre que se mueve por la plataforma). Dicho de otra forma, la imagen de segundo plano proporciona un escenario en el cual los elementos de primer plano pueden entrar, interactuar entre sí, y salir del mismo. La imagen de segundo plano puede incluir un valor de color para cada píxel en la imagen de segundo plano. En una realización, la imagen de segundo plano puede obtenerse mediante el muestreo de valores de color para un píxel dado a lo largo de un número de tramas. Así mismo, a medida que se reciben nuevas tramas, pueden actualizarse elementos de la imagen de segundo plano basándose en una información adicional incluida en cada trama sucesiva. Por lo general, qué píxeles son partes del segundo plano o el primer plano puede determinarse para cada trama en una secuencia de tramas de vídeo, y los elementos de primer plano pueden identificarse mediante la comparación de la imagen de segundo plano con los valores de color de píxel en una trama dada. Una vez que los píxeles de primer plano se han identificado, puede aplicarse una máscara a la trama, recortando de forma efectiva los píxeles que son parte del segundo plano con respecto a una imagen, dejando solo una o más manchas de píxeles de primer plano en la imagen. Por ejemplo, podrían aplicarse máscaras a una trama de tal modo que cada píxel de primer plano se representa como blanco y cada píxel de segundo plano se representa como negro. La imagen en blanco y negro resultante (que se representa como una matriz bidimensional) puede proporcionarse a elementos subsiguientes del sistema de reconocimiento de conducta. En una realización, el sistema de ordenador 115 puede proveerse con modelos iniciales de una imagen de segundo plano para una escena dada.

En la etapa 225, una imagen de primer plano asociada con una trama dada puede analizarse para identificar un conjunto de manchas (es decir, un grupo de píxeles relacionados) mediante la segmentación de la imagen de primer plano en unos objetivos de interés. Dicho de otra forma, el sistema puede configurarse para aislar manchas diferenciadas dentro de la imagen de primer plano, en el que es probable que cada mancha represente un objeto en primer plano diferente dentro de la trama (por ejemplo, un coche, un hombre; una maleta, y similares). Para cada mancha de primer plano, puede inicializarse un modelo de búsqueda cuando una mancha de primer plano se identifica inicialmente. El modelo de búsqueda se usa para capturar una posición de una mancha dentro del esquema, identificar qué píxeles están incluidos como parte de la mancha, y almacenar una diversidad de metadatos con respecto a la conducta observada de la mancha de trama a trama. Además, el modelo de búsqueda puede usarse por un módulo de rastreo para predecir, encontrar y rastrear los movimientos de un objeto correspondiente

de trama a trama. A medida que se reciben tramas sucesivas, el modelo de búsqueda se actualiza mientras la mancha de primer plano continúe encontrándose presente a través de tramas de vídeo sucesivas. Tales actualizaciones pueden realizarse con cada trama de vídeo adicional, de manera periódica, a medida que se recibe nueva información que permita que se refine el modelo de búsqueda, según sea necesario, o similar.

5 El modelo de búsqueda puede implementarse en una diversidad de formas. Por ejemplo, en una realización, el modelo de búsqueda puede ser un modelo de apariencia configurado para capturar un número de características acerca de un objeto en primer plano dado, incluyendo qué píxeles se consideran parte de ese objeto en primer plano. El modelo de apariencia de un objeto dado puede actualizarse a continuación, basándose en los píxeles que representan ese objeto de trama a trama. En otra realización, el modelo de búsqueda puede ser un rectángulo circunscrito mínimo para abarcar un objeto. A pesar de que se calcula más rápidamente, un rectángulo mínimamente circunscrito incluye píxeles como parte de la mancha que son, de hecho, parte del segundo plano. Sin embargo, para algunos tipos de análisis, este enfoque puede ser efectivo. Estos modelos de búsqueda se describen en lo sucesivo con mayor detalle. En la etapa 230, los modelos de búsqueda se usan para rastrear los movimientos de los objetos en primer plano cuando estos se mueven por la escena de trama a trama. Es decir, una vez que un objeto se ha identificado en una primera trama y un modelo de apariencia (y / o recuadro circunscrito) se ha generado para ese objeto, el modelo de búsqueda puede usarse para identificar y rastrear ese objeto en tramas subsiguientes, basándose en el modelo de apariencia (y / o recuadro circunscrito), hasta que ese objeto en primer plano sale de la escena. El modelo de búsqueda puede usarse para identificar un objeto dentro de las tramas de vídeo después de que el objeto, por ejemplo, cambios la ubicación o posición. Por lo tanto, se determinan diferentes tipos de información con respecto a los mismos objetos (por ejemplo, características cinemáticas del objeto, orientación, dirección de movimiento, y así sucesivamente) a medida que un objeto de este tipo se mueve a través de la escena.

En la etapa 235, el sistema de reconocimiento de conducta intenta clasificar las manchas de primer plano como una de un número discreto de clasificaciones. Por ejemplo, en una realización, el sistema de reconocimiento de conducta puede configurarse para clasificar cada objeto en primer plano como uno de un “ser humano”, un “vehículo”, “otro”, o “desconocido”. Por supuesto, pueden usarse más clasificaciones y además, las clasificaciones pueden adaptarse para adecuarse a las necesidades de un caso individual. Por ejemplo, un sistema de reconocimiento de conducta que recibe imágenes de vídeo de una cinta transportadora de equipaje podría clasificar los objetos sobre la cinta como diferentes tipos / tamaños de equipaje. Después de clasificar un objeto en primer plano, pueden hacerse estimaciones adicionales con respecto a tal objeto, por ejemplo, se estiman la postura (por ejemplo, orientación, actitud, y similares), la ubicación (por ejemplo, la ubicación dentro de una escena representada por las imágenes de vídeo, la ubicación en relación con otros objetos de interés, y similares), y el movimiento (por ejemplo, trayectoria, rapidez, dirección, y similares) del objeto. Esta información puede usarse por el motor de aprendizaje automático 140 para caracterizar determinadas conductas como normales o anómalas, basándose en observaciones pasadas de objetos similares (por ejemplo, otros objetos clasificados como seres humanos).

En la etapa 240, los resultados de etapas previas (por ejemplo, los resultados de rastreo, los datos de imagen de segundo plano / primer plano, los resultados de clasificación, y así sucesivamente) se combinan y se analizan para crear un mapa de una escena representada por las tramas de vídeo. En una realización, la escena se segmenta en regiones espacialmente separadas, estando definido cada segmento por un conjunto de píxeles. Las regiones se clasifican de acuerdo con la profundidad z (es decir, qué segmento se encuentra más cerca y qué segmento se encuentra más lejos de un dispositivo de captura de vídeo) y se etiquetan de forma opcional (por ejemplo, como naturales, fabricadas por el hombre, etc.). En la etapa 245, se crean representaciones semánticas de los movimientos de los objetos. Dicho de otra forma, se crean representaciones simbólicas de los movimientos y / o las acciones de los objetos rastreados (por ejemplo, “el coche aparca”, “el coche se detiene”, “la persona se inclina”, “la persona desaparece”, y así sucesivamente). En la etapa 250, las representaciones semánticas se analizan en busca de patrones reconocibles.

Las representaciones semánticas resultantes, el mapa con comentarios de una escena, y los resultados de clasificación se analizan en la etapa 255. El sistema de reconocimiento de conducta analiza tales resultados para aprender patrones de conducta, generaliza basándose en observaciones, y aprende mediante la realización de analogías. Esto también permite que el sistema de reconocimiento de conducta determine y / o aprenda qué tipo de conducta es normal y qué tipo de conducta es anormal. Es decir, el motor de aprendizaje automático puede configurarse para identificar patrones reconocibles, evaluar nuevas conductas para un objeto dado, reforzar o modificar los patrones de conductas aprendidos acerca de un objeto dado, etc.

En la etapa 260, los resultados de las etapas previas se analizan de forma opcional para una conducta reconocida. Adicionalmente, el sistema de reconocimiento de conducta puede configurarse para realizar una acción especificada en respuesta a reconocer la aparición de un suceso dado. Por ejemplo, basándose en los resultados de etapas previas, el sistema de reconocimiento de conducta puede emitir una alerta cuando un objeto en primer plano clasificado como un ser humano toma parte en una conducta inusual. Además, que una cierta conducta sea “inusual” puede estar basado en lo que el motor de aprendizaje ha “aprendido” que es una conducta “normal” para seres humanos en una escena dada. En una realización, se emiten alertas solo si se ha determinado una conducta anormal (por ejemplo, una alerta que indica que una persona dejó una bolsa sin vigilancia en una estación de metro). En otra realización, se emiten alertas para indicar que están teniendo lugar sucesos normales en la escena

(por ejemplo, una alerta que indica que un coche aparcó). El procedimiento concluye con la etapa 275.

Debe observarse que no es necesario realizar la totalidad de las etapas que se han descrito en lo que antecede en el orden indicado. Además, no todas las etapas que se describen son necesarias para que funcione el procedimiento descrito. Qué etapas deberían usarse, en qué orden deberían realizarse las etapas, y si algunas etapas deberían repetirse más a menudo que otras etapas se determina, basándose en, por ejemplo, las necesidades de un usuario particular, las cualidades específicas de un entorno observado, y así sucesivamente.

Las figuras 3 a 6 ilustran diferentes componentes del motor de visión de ordenador 135 que se ilustra en la figura 1, de acuerdo con una realización de la presente invención. De manera específica, la figura 3 ilustra componentes de un módulo de segundo plano - primer plano 300. El módulo de segundo plano - primer plano 300 usa características en cada trama de vídeo para identificar qué píxeles pertenecen a una imagen de segundo plano y cuales pertenecen a una imagen de primer plano. En una realización, se analizan tramas de vídeo para clasificar cada píxel como que visualiza parte de la imagen de segundo plano para la escena (y esa trama) o que visualiza parte de una imagen de primer plano para esa trama.

Por lo general, píxeles que no cambian de color con el tiempo se consideran parte de la imagen de segundo plano. Mediante el muestreo del valor de color de un píxel con el tiempo, la presencia de un objeto en primer plano en algunas tramas puede eliminarse. Además, debido a que la imagen de segundo plano puede actualizarse de manera dinámica, la imagen de segundo plano puede compensar los cambios en luz y sombra. De forma similar, se supone que los píxeles que cambian de color, en relación con la imagen de segundo plano, están visualizando un objeto en primer plano. Dicho de otra forma, los movimientos de los objetos en primer plano en una escena se determinan basándose en diferencias entre valores de color de píxel en tramas de vídeo sucesivas. En general, una imagen de segundo plano puede concebirse como una trama de vídeo de píxeles que tiene los objetos en primer plano recortados. Las imágenes de primer plano pueden concebirse como píxeles que ocuyen el segundo plano. Como alternativa, puede usarse solo una imagen de primer plano. Tal imagen de primer plano puede concebirse como una trama de vídeo transparente con parches de los píxeles de primer plano. Debería observarse, que a pesar de que dos tramas consecutivas pueden ser suficientes para rastrear un objeto en primer plano dado, comparar múltiples tramas consecutivas proporciona resultados más precisos cuando se determina la imagen de segundo plano para una escena dada.

También debe observarse, que un píxel originalmente determinado como un píxel de segundo plano (en una trama) puede volverse un píxel de primer plano (en otra trama) y viceversa. Por ejemplo, si el valor de color de un píxel en el segundo plano comienza a cambiar, puede ser apropiado volver a clasificar este como un píxel de primer plano (por ejemplo, un coche aparcado en una zona de aparcamiento durante un periodo de tiempo prolongado empieza a moverse). De forma similar, un píxel cambiante podría volverse estático, por lo tanto podría ser necesario volver a calificar tal píxel como un píxel de segundo plano (por ejemplo, una papelera se ha llevado a una estación de metro para un uso permanente). No obstante, para evitar una reclasificación de píxeles innecesaria y para mejorar la interpretación de lo que está incluido en las imágenes de segundo plano y de primer plano, en una realización, el sistema de reconocimiento de conducta puede clasificar píxeles como parte de un segundo plano a corto plazo (STBG), un primer plano a corto plazo (STFG), un segundo plano a largo plazo (LTBG) y un primer plano a largo plazo (LTFG). STBG y STFG se almacenan en memoria durante un corto periodo de tiempo (por ejemplo, segundos o menos), mientras que LTBG y LTFG se almacenan en memoria para periodos de tiempo más prolongados (por ejemplo, minutos). Determinar los píxeles que van a ser STBG / STFG al principio, y a continuación interpretar solo los píxeles de calificación como LTBG / LTFG permite una determinación más precisa de qué píxeles son parte de la imagen de segundo plano / primer plano. Por supuesto, los periodos de tiempo pueden ajustarse de acuerdo con los sucesos que tienen lugar dentro de una escena particular.

La figura 3 ilustra componentes del módulo de segundo plano - primer plano 300 que pueden usarse para generar imágenes de segundo plano y de primer plano para una trama de vídeo, de acuerdo con una realización de la invención. Inicialmente, se reciben tramas de vídeo por un módulo de entrenamiento de segundo plano 305. El módulo de segundo plano - primer plano 300 puede entrenarse usando una secuencia inicial de tramas. El entrenamiento permite que el módulo de segundo plano - primer plano 300 construya una imagen de segundo plano de la escena representada en las tramas de vídeo adquiridas. El procedimiento de entrenamiento puede tener lugar durante una fase de inicialización del sistema; a saber, antes de que se haya determinado una imagen de segundo plano de la escena.

El módulo de compensación de escenas oscuras 310 puede procesar valores de píxel para compensar unas condiciones de iluminación baja u oscura en porciones de la escena. Adicionalmente, el módulo de compensación de escenas oscuras 310 puede configurarse para proporcionar las tramas de vídeo procesadas a un módulo de STFG / STBG 315 y un módulo de LTBG / LTBG 320. El módulo de STFG / STBG 315 puede configurarse para identificar píxeles de STFG y de STBG dentro de una trama dada y proporcionar esta información a un módulo de FG obsoleto 325 y un módulo de compensación de iluminación 335, respectivamente. El módulo de LTFG / LTBG 320 puede configurarse para identificar píxeles de LTFG y de LTBG y, de forma similar al módulo de STFG / STBG 315, proporcionar esta información al módulo de FG obsoleto 325 y el módulo de compensación de iluminación 335, respectivamente. El módulo de FG obsoleto 325 identifica píxeles de primer plano obsoletos y proporciona los resultados a un módulo de BG de actualización 330. Un píxel puede volverse "obsoleto" cuando, la determinación de

BG / FG es obsolescente y es necesario volver a evaluarla. Una vez que se ha recibido, el módulo de compensación de iluminación 335 puede ajustar de manera dinámica el procesamiento para los cambios en la iluminación (por ejemplo, la intensificación de brillo / oscurecimiento de una escena debido a nubes que oscurecen el sol, o ajustes a las fuentes de luz artificial), y el módulo de compensación de escenas oscuras 310 proporcionará de manera dinámica un procesamiento especial en el límite de regiones extremadamente oscuras y / o condiciones de baja luz. El módulo de BG de actualización 330 actualiza un modelo de imágenes de segundo plano y transfiere los resultados al módulo de compensación de iluminación 335, que a su vez, después del procesamiento de la totalidad de los resultados recibidos, proporciona los resultados procesados al módulo de LTFG / LTBG.

Por lo tanto, de forma colectiva, el módulo de segundo plano - primer plano 300 determina un conjunto de imágenes de segundo plano y de primer plano y / o modelos de segundo plano y de primer plano para su uso por otros componentes del sistema de reconocimiento de conducta. Los modelos de segundo plano y de primer plano distinguen entre píxeles que son parte del segundo plano de la escena (es decir, parte del escenario) y píxeles que visualizan objetos en primer plano (es decir, elementos que realizan una cierta acción sobre el escenario). Debería observarse que a pesar de que en la descripción anterior del módulo de segundo plano - primer plano 300 las referencias se hacen solo a una imagen de segundo plano, como alternativa, el módulo de segundo plano - primer plano 300 puede emplear múltiples imágenes de segundo plano (por ejemplo, la escena de la trama de imagen podría dividirse en varias zonas de segundo plano para una identificación de segundo plano más precisa).

En una realización, la imagen / modelo de segundo plano puede incluir una información adicional, tal como colores de píxel. Además, la imagen / modelo de primer plano por lo general incluye características de píxel adicionales, tales como color. No obstante, mantener o recopilar tal información podría omitirse (por ejemplo, para conservar recursos en un entorno en el que conocer los colores no mejora de manera significativa la distinción entre objetos de interés, por ejemplo una línea de transporte que transporta objetos en su mayor parte del mismo color, o uno similar).

La figura 4 ilustra un módulo de objetos en primer plano 400 configurado para identificar objetos visualizados en las imágenes de primer plano de una escena, de acuerdo con una realización de la invención. En general, el módulo de objetos en primer plano 400 puede configurarse para recibir las imágenes de primer plano producidas por el módulo de segundo plano - primer plano 300 para una trama dada, construir / actualizar modelos de búsqueda para las imágenes de primer plano, e intentar rastrear los movimientos de un objeto visualizado en las imágenes de primer plano cuando ese objeto se mueve por la escena de trama a trama.

Tal como se ilustra en la figura 4, el módulo de objetos en primer plano 400 incluye un módulo de detección de manchas 405, un módulo de construcción / actualización 410, un módulo de rastreo 420 y 1-M modelos de búsqueda, del modelo de búsqueda 1 (415_1), el modelo de búsqueda 2 (415_2), al modelo de búsqueda M (415_M). En una realización, el módulo de detección de manchas 405 puede configurarse para analizar imágenes de primer plano para detectar grupos de píxeles relacionados, a los que se hace referencia como las manchas de primer plano, en las que es probable que cada grupo de píxeles de este tipo represente un objeto diferenciado dentro de la escena. Adicionalmente, a cada mancha de primer plano detectada se le asigna un número de identificación de rastreo. Las manchas de primer plano se usan por el módulo de construcción / actualización 410 para construir / actualizar los modelos de búsqueda $415_1 - 415_M$, en el que los modelos de búsqueda ya existentes se han construido o actualizado para manchas identificadas en tramas de vídeo previas. En una realización, para actualizar los modelos de búsqueda $415_1 - 415_M$, el módulo de construcción / actualización 410 también usa resultados que se generan por el módulo de rastreo 420. Si una mancha actualmente detectada no tiene modelo de búsqueda respectivo, se construye (se crea) tal modelo de búsqueda.

En cualquier instante dado, el módulo de objetos en primer plano 400 puede incluir múltiples modelos de búsqueda, representando cada uno una mancha de primer plano diferente. El número de modelos de búsqueda puede depender de cuántas manchas de primer plano se identifican por el módulo de detección de manchas 405 dentro de una imagen de primer plano. En una realización, los modelos de búsqueda pueden configurarse con capacidades predictivas con respecto a lo que las manchas de primer plano pueden hacer en tramas de vídeo subsiguientes. Por ejemplo, el modelo de búsqueda asociado con una mancha de primer plano dada puede incluir una posición (y forma) futura esperada de esa mancha basándose en una posición, y unos datos cinemáticos, presentes. Además, cada modelo de búsqueda también puede incluir una diversidad de informaciones que se obtiene acerca de una mancha de primer plano dada (por ejemplo, texturas, colores, patrones, la posición de profundidad z dentro de una escena, tamaño, tasas de movimiento, cinemática y similares).

Además, diferentes tipos de modelos de búsqueda pueden usarse de acuerdo con los principios de la presente invención. Tal como se indica, un modelo de búsqueda puede usarse por el módulo de rastreo 420 para predecir, encontrar y rastrear los movimientos de un objeto correspondiente de trama a trama. En una realización, se usa un modelo de apariencia. El modelo de apariencia incluye píxeles que se usan para visualizar un objeto (por ejemplo, cuando una trama visualiza un ser humano en la imagen de primer plano, el modelo de apariencia incluiría en su mayor parte píxeles que perfilan el ser humano y píxeles en el interior del perfil). En otra realización, el modelo de búsqueda se implementa como un modelo basado en características, en el que el modelo basado en características representa píxeles en el interior de un rectángulo, tal como un rectángulo circunscrito mínimo que abarca un objeto (por ejemplo, cuando un objeto es un ser humano, el modelo basado en características podría incluir un rectángulo

circunscrito que abarca el ser humano). Como alternativa, el modelo basado en características puede incluir múltiples rectángulos circunscritos para un objeto dado, tal como rectángulos de tamaños mínimamente posibles, que abarcan diferentes regiones de ese objeto (por ejemplo, cuando la trama visualiza un ser humano, el modelo basado en características para tal objeto podría incluir varios rectángulos de tamaño mínimo en el que los rectángulos abarcan diferentes regiones del ser humano, tal como brazos, piernas, cabeza y torso).

Qué modelo de búsqueda se usa puede depender, por ejemplo, de un entorno que se está observando, de las preferencias de un usuario de un sistema de reconocimiento de conducta, y así sucesivamente. Por ejemplo, a pesar de que es probable que el modelo de apariencia proporcione un rastreo más preciso, el modelo basado en características puede conservar recursos, cuando, por ejemplo, las formas de los objetos de interés rastreados son simples (por ejemplo, una cinta transportadora de equipaje).

Tal como se ha mencionado en lo que antecede, el módulo de rastreo 420 usa los modelos de búsqueda 415 para rastrear los movimientos de los objetos correspondientes representados en una secuencia de vídeo de trama a trama cuando tales objetos se mueven por la escena. El módulo de rastreo 420 toma una mancha de primer plano detectada de una trama de vídeo actual y busca un modelo de búsqueda de una trama de vídeo previa que proporciona la coincidencia más aproximada con la mancha de primer plano. En una realización, para cada mancha de primer plano actualmente detectada, el módulo de rastreo 420 busca un modelo de búsqueda 415 que una distancia vectorial dimensional relativa entre el modelo de búsqueda y la mancha de primer plano es un mínimo global. De esta forma, el módulo de rastreo 420 puede rastrear las ubicaciones de cada objeto representado por uno de los modelos de búsqueda 415 de trama a trama. En una realización, el módulo de rastreo 420 usa información cinemática adquirida basándose en tramas de vídeo previas para estimar ubicaciones del modelo de búsqueda dentro de la trama de vídeo actual.

La figura 5 ilustra un módulo de estimador / identificador 500 de un motor de visión de ordenador, de acuerdo con una realización de la presente invención. En general, el estimador / identificador 500 recibe manchas de primer plano y modelos de búsqueda respectivos e intenta clasificar objetos en una trama de vídeo, tal como se representa por las manchas de primer plano, como miembros de categorías conocidas (clases). En una realización, si un objeto de interés se ha identificado a continuación el módulo de estimador / identificador 500 estima la postura, la ubicación y el movimiento del objeto de interés. El estimador / identificador 500 se entrena por lo general en numerosos ejemplos positivos y negativos que representan ejemplos de una clase dada. Además, un entrenamiento en línea puede usarse para actualizar el clasificador dinámicamente a la vez que se analiza vídeo trama a trama.

Como se muestra, el estimador / identificador 500 incluye un clasificador 505, de la clase 1 (510_1) a la clase N (510_N), y un identificador 515. El clasificador 505 intenta clasificar un objeto en primer plano como un miembro de una de las clases, de la clase 1 (510_1) a la clase N (520_N). Si es satisfactorio, también pueden determinarse datos estáticos (por ejemplo, tamaño, color, y similares) y datos cinemáticos (por ejemplo, rapidez, velocidad, dirección y similares) representativos del objeto clasificado a lo largo de un periodo de tiempo (por ejemplo, X - número de tramas) por el identificador 515. Para cada objeto identificado, el estimador / identificador 500 emite sucesos de contexto sin procesar que contienen las características estáticas y cinemáticas que se han descrito en lo que antecede del objeto de interés y observaciones de objeto conocidas que contienen características estáticas y cinemáticas de un miembro promedio de la clase del objeto identificado.

En una realización, el sistema emplea cuatro clasificadores: ser humano, vehículo, otro, y desconocido. Hasta que se determine una clase de objeto de interés, tal objeto se trata como un miembro de clase "desconocido". Cada clase contiene postura, datos estáticos y cinemáticos con respecto a un miembro promedio de la clase. En una realización, tales datos se actualizan de forma continua a medida que se clasifican e identifican más objetos de interés y se determinan y recopilan su postura, datos estáticos y cinemáticos. Debería observarse que, por lo general, el estimador / identificador 500 procesa información en tiempo real, de una forma trama a trama.

La figura 6 ilustra un procesador de contexto 600 de un motor de visión de ordenador 135, de acuerdo con una realización de la presente invención. En general, el procesador de contexto 600 combina resultados que se reciben a partir de otros componentes del motor de visión de ordenador 135, tal como el módulo de segundo plano - primer plano 300, el módulo de objetos en primer plano 400, y el módulo de estimador / identificador 500, para crear un mapa con comentarios de una escena capturada en las tramas de vídeo. En una realización, la escena se segmenta en regiones espacialmente separadas que se clasifican de acuerdo con la profundidad z de la escena y se etiquetan de forma opcional como que representa elementos de origen natural o fabricados por el hombre.

Como se muestra, el procesador de contexto 600 puede incluir un segmentador en regiones 605 para descomponer la escena en unas áreas (regiones) más pequeñas, un secuenciador de regiones 610 para definir relaciones entre las regiones (por ejemplo, como más cerca / más lejos de un dispositivo de captura de vídeo en relación entre sí), y un generador de mapas de escena 615 para generar el mapa con comentarios. En una realización, el procesador de contexto 600 usa información con respecto a los movimientos (tal como trayectorias) y las ubicaciones de los objetos de interés rastreados para generar el mapa con comentarios.

Las figuras 7 y 8 ilustran diferentes componentes del motor de aprendizaje automático 140 que se ilustra en la figura 1. De manera específica, la figura 7 ilustra componentes de un módulo de análisis semántico 700 y la figura 8 ilustra

componentes de un módulo de percepción 800, de acuerdo con una realización de la presente invención. En general, el módulo semántico 700 crea representaciones semánticas (es decir, representaciones simbólicas) de los movimientos y las acciones de los objetos rastreados. La representación semántica proporciona una manera formal de describir lo que se cree que está sucediendo en la escena basándose en los movimientos de un objeto rastreado particular (y en última instancia, basándose en los cambios en los valores de color de píxel de trama a trama). Se usa una gramática de lenguaje formal (por ejemplo, nombres y verbos) para describir sucesos en la escena (por ejemplo, “el coche aparca”, “la persona aparece”, y similares).

Subsiguientemente, las representaciones semánticas se analizan en busca de patrones reconocibles y los resultados se proporcionan a un módulo de percepción 800 que se ilustra en la figura 8. En una realización, el módulo semántico 700 también construye un mapa simbólico de la escena, incluyendo diferentes aspectos de los sucesos que tienen lugar en la escena, tal como representaciones simbólicas de trayectorias de los objetos en la escena. En una realización, el mapa simbólico también puede incluir una distribución de frecuencia (por ejemplo, datos con respecto a con qué frecuencia y dónde están presentes determinadas clases o tipos de objetos en la escena).

Tal como se muestra en la figura 7, el módulo semántico 700 incluye una memoria sensorial 710, un módulo de análisis semántico latente (LSA) 715, un módulo de sucesos primitivos 725, un módulo de división de espacio de fase 730, un módulo de análisis semántico latente incremental (iLSA) 735, y un módulo de lenguaje formal 740. La memoria sensorial 710 adquiere información proporcionada para el módulo semántico 700 y almacena esta información para un uso subsiguiente por el módulo de sucesos primitivos 725 y el módulo de división de espacio de fase 730. En una realización, la memoria sensorial 710 identifica qué información debería proporcionarse para un análisis adicional al módulo de sucesos primitivos 725 y el módulo de división de espacio de fase 730.

El módulo de detección de sucesos primitivos 725 puede configurarse para identificar la aparición de sucesos primitivos (por ejemplo, el coche se detiene, invierte el sentido, desaparece, aparece; la persona se inclina, cae; intercambio, y similares) en la memoria sensorial 710. Los sucesos primitivos por lo general reflejan los cambios en las características cinemáticas de los objetos rastreados. Por lo tanto, una vez que un objeto se ha clasificado como un “coche”, el módulo de detección de sucesos primitivos 725 puede evaluar datos con respecto al coche para identificar diferentes sucesos conductuales a medida que estos tienen lugar. En una realización, los sucesos primitivos están predefinidos (por ejemplo, para un entorno específico en el que se usa el sistema de reconocimiento de conducta con autoaprendizaje). En otra realización, solo algunos de los sucesos primitivos están predefinidos (por ejemplo, aparcamiento, giro, caída), mientras que otros sucesos primitivos se aprenden con el tiempo (por ejemplo, objetos de una determinada clase pueden encontrarse en un punto específico de la escena).

El módulo de división de espacio de fase 730 determina información con respecto a la posición geométrica que tiene velocidad de los objetos en la escena. Por consiguiente, el módulo de sucesos primitivos 725 y el módulo de división de espacio de fase 730 permiten que el módulo semántico 700 analice datos de dos maneras diferenciadas. En base a los resultados del módulo de sucesos primitivos 725 y el módulo de división de espacio de fase 730, el LSA 715 y el iLSA 735 construyen / actualizan un modelo de la escena, en el que el modelo incluye los objetos de interés.

LSA 715 es en general un módulo de entrenamiento inicial del módulo semántico 700. LSA reúne datos a lo largo de un periodo de tiempo hasta que el LSA 715 genera resultados de suficiente peso estadístico. Dicho de otra forma, LSA 715 aprende la distribución básica de la escena, mientras que iLSA 735 actualiza de manera incremental una distribución de este tipo. Debe observarse que iLSA 735 es lo bastante flexible para manejar los cambios en los patrones de conducta que tienen lugar en la escena. El módulo de aprendizaje de lenguaje formal 740 usa los datos que se generan por el iLSA 735 para crear las representaciones semánticas (la representación simbólica de lo que está sucediendo en la escena) y proporciona las representaciones semánticas al módulo de percepción 800 para aprender lo que significan las representaciones semánticas creadas.

La figura 8 ilustra un módulo de percepción de un motor de aprendizaje automático, de acuerdo con una realización de la invención. El módulo de percepción 800 puede configurarse para procesar los resultados que se generan por al menos algunos de los componentes de la visión de ordenador 135 y el motor de aprendizaje automático 140 (por ejemplo, el módulo de estimador / identificador 500, el procesador de contexto 600, el módulo semántico 700, etc.). En general, el módulo de percepción 800 aprende patrones, generaliza basándose en observaciones, y aprende mediante la realización de analogías.

Como se muestra en la figura 8, el módulo de percepción 800 puede incluir una memoria asociativa perceptiva 805, un programador 810, un espacio de trabajo 815, una memoria episódica 820, y una memoria a largo plazo 825. El espacio de trabajo 815 proporciona una región de memoria que refleja qué información se está evaluando en la actualidad por el motor de aprendizaje automático 140. Es decir, el espacio de trabajo 815 almacena qué elementos de datos tienen en la actualidad la “atención” del entorno de aprendizaje automático 140. Tal como se describe en lo sucesivo, los datos en el espacio de trabajo 815 pueden incluir una colección de perceptos (describiendo cada uno un suceso) y fragmentos de código especializado (la memoria asociativa perceptiva 805 recoge datos proporcionados al módulo de percepción 800 y almacena tales datos como perceptos. Cada percepto puede proporcionar datos que describen algo que tuvo lugar en el vídeo, tal como un suceso primitivo. La memoria asociativa perceptiva 805 proporciona perceptos y / o fragmentos de código especializado al espacio de trabajo 815.

Un fragmento de código especializado proporciona un fragmento de código ejecutable, que describe y / o busca relaciones entre diferentes perceptos. Dicho de otra forma, un fragmento de código especializado resume reglas para determinar una conducta / suceso específico (por ejemplo, suceso de aparcamiento), en el que la conducta / suceso implica uno o más perceptos. Cada fragmento de código especializado puede configurarse para tomar un conjunto de preceptos de entrada y procesarlos de una manera particular. Por ejemplo, un fragmento de código especializado puede tomar un conjunto de perceptos de entrada y evaluarlos para determinar si ha tenido lugar un suceso particular (por ejemplo, un coche aparcando). Usando el ejemplo de un coche aparcando, el percepto puede actualizar la memoria episódica 820 con información acerca de qué coche, el color del coche, en dónde el coche aparcó, etc. Además, la información acerca de este suceso primitivo detectado puede usarse para actualizar la definición del suceso primitivo en la memoria a largo plazo 825. Aún más, fragmentos de código especializado que reconocen anomalías se emplean por el módulo de percepción 800. Tales fragmentos de código especializado acceden a perceptos y si un determinado percepto no se correlaciona estadísticamente con datos estadísticos previamente acumulados, puede identificarse un suceso anormal.

En una realización, los fragmentos de código especializado están completamente pre-escritos. En otra realización, al menos algunos fragmentos de código especializado no están completamente pre-escritos, sino que en su lugar, se generan con el tiempo. Por ejemplo, un fragmento de código especializado que describe una conducta normal para un determinado percepto(s) puede modificarse / generarse a sí mismo basándose en datos acumulados que describen sucesos observados correspondientes.

El programador 810 determina qué fragmento de código especializado es necesario activar en cualquier instante dado. Por ejemplo, el programador 810 puede buscar identificar una coincidencia entre perceptos ubicados en el espacio de trabajo 815 y los fragmentos de código especializado. Cuando se encuentra disponible un conjunto apropiado de entradas requeridas para un fragmento de código especializado dado (por ejemplo, un conjunto de preceptos), ese fragmento de código especializado puede ubicarse en el espacio de trabajo 815 e invocarse. Cuando múltiples fragmentos de código especializado se encuentran disponibles para su activación, la determinación de cuándo y qué fragmento de código especializado activar puede ser aleatoria. No obstante, en una realización, determinados fragmentos de código especializado configurados tendrán prioridad frente a otros (por ejemplo, un fragmento de código especializado que define una determinada conducta anormal). En cada instante dado, numerosos fragmentos de código especializado pueden activarse por el programador 810 dentro del espacio de trabajo 815.

El módulo de percepción 800 también usa la memoria episódica 820 y la memoria a largo plazo 825 para capturar datos tanto a corto plazo como a largo plazo con respecto a sucesos primitivos. La memoria episódica 820 es una memoria a corto plazo para almacenar perceptos recientes. Por ejemplo, un percepto que se ha cambiado recientemente se encuentra en la memoria episódica 820. Los perceptos se ubican en la memoria episódica 820 del espacio de trabajo 815. Al mismo tiempo, el espacio de trabajo 815 puede usar los perceptos almacenados en la memoria episódica 820 para hacer que coincidan con los fragmentos de código especializado respectivos.

Por lo general, al menos algunos perceptos emigran de la memoria episódica 820 a la memoria a largo plazo 825. No obstante, no todos los fragmentos de datos ubicados en la memoria episódica 820 emigran a la memoria a largo plazo 825. Algunos datos decaen de la memoria episódica 820 sin alcanzar nunca la memoria a largo plazo 825 (por ejemplo, datos que describen un suceso que ocurre una sola vez que no se ha determinado como anormal).

Al mismo tiempo, los aspectos de ese suceso pueden usarse para reforzar la información en la memoria a largo plazo 825 (por ejemplo, aspectos de cómo, dónde, y por cuánto tiempo un coche aparcó en un espacio de aparcamiento). Por lo tanto, la memoria a largo plazo 825 puede usarse para construir y acumular patrones generales de conducta dentro de una escena dada. En una realización, los patrones de conducta almacenados en la memoria episódica 820 y los patrones de conducta que han adquirido suficiente peso estadístico se mueven a la memoria a largo plazo 825 como los patrones generales de conducta. No obstante, no todos los datos ubicados en la memoria a largo plazo 825 permanecen ahí. Algunos datos finalmente decaen (por ejemplo, detalles específicos). Por ejemplo, si se han aparcado varios coches de diferentes colores en el mismo lugar a lo largo de un periodo de tiempo, puede aprenderse un patrón general de un coche que puede aparcar en ese lugar específico y colocarse en la memoria a largo plazo 825. No obstante, los detalles con respecto a los coches previamente aparcados, tal como sus colores, decaerían de la memoria a largo plazo 825 después de un cierto periodo de tiempo.

En una realización, el espacio de trabajo 815 usa los patrones generales de conducta encontrados en la memoria a largo plazo 825 para determinar sucesos que tienen lugar en la escena. Una vez que un suceso se ha reconocido, se genera la información que indica que el suceso reconocido se ha identificado. Tal información se usa subsiguientemente para generar alertas. Mientras que en una realización, solo se emiten alertas con respecto a una conducta anormal identificada (por ejemplo, asalto), en otra realización, se emiten así mismo alertas que se describen como normales (por ejemplo, coche aparcado).

Las figuras 9A-9C ilustran un escenario que tiene lugar place en una estación de metro 900 en el que un sistema de reconocimiento de conducta detecta una conducta anormal y emite una alerta, de acuerdo con una realización de la presente invención. Tal como se muestra, una cámara de vídeo estacionaria 915 captura sucesos que tienen lugar en la estación de metro 900 y proporciona unas imágenes de vídeo que representan los sucesos para el sistema de

reconocimiento de conducta. Tal como se ilustra en las figuras 9A-9C, la cámara de vídeo 915 captura unas imágenes de vídeo de un hombre 905 que porta una bolsa 910 a la vez que se aproxima a la papelería 920 (figura 9A), pone la bolsa 910 en el suelo junto a la papelería 920 (figura 9B), y deja la bolsa 910 atrás (figura 9C). En base al aprendizaje a partir de la observación de seres humanos, entrar en la estación de metro 900, el acto de dejar "otro" objeto (es decir, la bolsa) portado por un objeto clasificado como un ser humano puede identificarse como anormal, y por consiguiente, el sistema de reconocimiento de conducta puede emitir una alerta para indicar la aparición de un suceso de este tipo.

De acuerdo con los principios que se han analizado en lo que antecede, el sistema de reconocimiento de conducta trata los píxeles que visualizan la papelería 920 estacionaria como una parte de una imagen de segundo plano, sin identificar de manera específica la papelería 920 como una papelería. Por el contrario, el sistema de reconocimiento de conducta trata tanto al hombre 905 como a la bolsa 910 como imagen o imágenes de primer plano. Inicialmente (figura 9A), el sistema de reconocimiento de conducta con autoaprendizaje puede considerar al hombre 905 y a la bolsa 910 como una mancha de primer plano. No obstante, cuando el hombre 905 pone la bolsa 910 en el suelo (figuras 9B-9C), el hombre y la bolsa 910 se vuelven partes de manchas de primer plano independientes. A pesar de que en una realización, cuando el hombre 905 recoge la bolsa 910 sus manchas de primer plano respectivas se fundirían en unas nuevas manchas de primer plano, en otra realización, el hombre 905 y la bolsa 910 siguen tratándose como dos manchas de primer plano diferenciadas. En aún otra realización, se considera que el hombre 905 y la bolsa 910 son manchas de primer plano independientes desde el principio (figuras 9A).

Tanto para el hombre 905 como la bolsa 910 el sistema de reconocimiento de conducta construye y actualiza modelos de búsqueda para rastrear estos objetos trama a trama. Además, el sistema de reconocimiento de conducta clasifica el hombre 905 como un "ser humano" y la bolsa 910 como "otro" (como alternativa como una "bolsa"), recopila información acerca de los mismos, y predice sus acciones basándose en una conducta previamente aprendida de personas y bolsas en la estación de metro. Debido a que dejar atrás una bolsa no está asociado con una conducta aprendida normal, el sistema de reconocimiento de conducta identifica tal conducta como anormal y emite una alerta. Como alternativa, tal conducta puede identificarse como anormal debido a que el sistema ha aprendido previamente que la situación de dejar atrás una bolsa indica una conducta anormal.

A pesar de que lo anterior se refiere a realizaciones de la presente invención, pueden concebirse otras, y adicionales, realizaciones de la invención sin apartarse del ámbito básico de la misma, y el alcance de la misma está determinado por las reivindicaciones que siguen.

REIVINDICACIONES

1. Un procedimiento para procesar un flujo de tramas de vídeo que registra sucesos dentro de una escena, comprendiendo el procedimiento:
 - 5 recibir una primera trama del flujo (210, 215), en el que la primera trama incluye datos para una pluralidad de píxeles incluidos en la trama; identificar uno o más grupos de píxeles en la primera trama, en el que cada grupo representa un objeto dentro de la escena (225); generar un modelo de búsqueda que almacena una o más características asociadas con cada objeto identificado;
 - 10 clasificar cada uno de los objetos usando un clasificador entrenado (235); rastrear, en una segunda trama, cada uno de los objetos identificados en la primera trama usando el modelo de búsqueda (230); suministrar la primera trama, la segunda trama y las clasificaciones de objetos a un motor de aprendizaje automático; y
 - 15 generar, por el motor de aprendizaje automático, una o más representaciones semánticas de conducta en la que toman parte los objetos en la escena a lo largo de una pluralidad de tramas (245), en el que el motor de aprendizaje automático está configurado para aprender patrones de conducta observada en la escena a lo largo de la pluralidad de tramas (255) y para identificar apariciones de los patrones de conducta en la que toman parte los objetos clasificados (260).
- 20 2. El procedimiento de la reivindicación 1, que comprende además emitir al menos una alerta que indica una aparición de uno de los patrones identificados de conducta por uno de los objetos rastreados.
3. El procedimiento de la reivindicación 1, en el que cada modelo de búsqueda es generado como uno de un modelo de apariencia y un modelo basado en características.
4. El procedimiento de la reivindicación 1, en el que la etapa de rastrear, en la segunda trama, cada uno de los
 - 25 objetos identificados en la primera trama usando el modelo de búsqueda comprende:
 - ubicar los objetos identificados dentro de la segunda trama; y
 - actualizar el modelo de búsqueda respectivo para cada objeto identificado.
5. El procedimiento de la reivindicación 1, en el que el clasificador entrenado está configurado para clasificar cada objeto como uno de un ser humano, un coche, u otro.
- 30 6. El procedimiento de la reivindicación 1, en el que la etapa de identificar uno o más grupos de píxeles en la primera trama comprende:
 - identificar al menos un grupo de píxeles que representan una región de primer plano de la primera trama y al menos un grupo de píxeles que representan una región de segundo plano de la primera trama;
 - 35 segmentar regiones de primer plano en manchas de primer plano, en el que cada mancha de primer plano representa un objeto representado en la primera trama; y
 - actualizar una imagen de segundo plano de la escena basándose en las regiones de segundo plano identificadas en la primera trama.
7. El procedimiento de la reivindicación 6, que comprende además:
 - 40 actualizar un mapa con comentarios de la escena representada por el flujo de vídeo usando los resultados de las etapas de generar un modelo de búsqueda que almacena una o más características asociadas con cada objeto identificado;
 - clasificar cada uno de los objetos usando un clasificador entrenado; y
 - 45 rastrear, en una segunda trama, cada uno de los objetos identificados en la primera trama usando el modelo de búsqueda, en el que el mapa con comentarios describe una geometría tridimensional de la escena que incluye una posición tridimensional estimada de los objetos identificados y una posición tridimensional estimada de una pluralidad de objetos representados en la imagen de segundo plano de la escena, y en el que la etapa de construir representaciones semánticas comprende además analizar las representaciones semánticas construidas para patrones de conducta reconocibles usando análisis semántico latente.
8. Un sistema, que comprende:
 - 50 una fuente de entrada de vídeo (105); un procesador (120); y una memoria (130) que almacena:
 - un motor de visión de ordenador (135), en el que el motor de visión de ordenador está configurado para:

- recibir, a partir de la fuente de entrada de vídeo, una primera trama de un flujo de vídeo, en el que la primera trama incluye datos para una pluralidad de píxeles incluidos en la trama, identificar uno o más grupos de píxeles en la primera trama, en el que cada grupo representa un objeto dentro de la escena (225),
- 5 generar un modelo de búsqueda que almacena una o más características asociadas con cada objeto identificado, clasificar cada uno de los objetos usando un clasificador entrenado (235), rastrear, en una segunda trama, cada uno de los objetos identificados en la primera trama usando el modelo de búsqueda (230), y
- 10 suministrar la primera trama, la segunda trama y las clasificaciones de objetos a un motor de aprendizaje automático; y
- el motor de aprendizaje automático, en el que el motor de aprendizaje automático está configurado para generar una o más representaciones semánticas de conducta en la que toman parte los objetos en la escena a lo largo de una pluralidad de tramas (245) y está configurado además para aprender patrones de conducta observada en la escena a lo largo de la pluralidad de tramas (255) y para identificar apariciones de los patrones de conducta en la que toman parte los objetos clasificados (260).
- 15
9. El sistema de la reivindicación 8, en el que el motor de aprendizaje automático está configurado además para emitir al menos una alerta que indica una aparición de uno de los patrones identificados de conducta por uno de los objetos rastreados.
- 20
10. El sistema de la reivindicación 8, en el que cada modelo de búsqueda es generado como uno de un modelo de apariencia y un modelo basado en características.
11. El sistema de la reivindicación 8, en el que la etapa de rastrear, en la segunda trama, cada uno de los objetos identificados en la primera trama usando el modelo de búsqueda comprende:
- 25 ubicar los objetos identificados dentro de la segunda trama; y actualizar el modelo de búsqueda respectivo para cada objeto identificado.
12. El sistema de la reivindicación 8, en el que el clasificador entrenado está configurado para clasificar cada objeto como uno de un ser humano, un coche, u otro.
13. El sistema de la reivindicación 12, en el que el clasificador entrenado está configurado además para estimar al menos uno de una postura, una ubicación, y un movimiento para al menos uno de los objetos clasificados, basándose en cambios en el grupo de píxeles que representan el objeto a lo largo de una pluralidad de tramas sucesivas.
- 30
14. El sistema de la reivindicación 8, en el que el motor de visión de ordenador está configurado para identificar el uno o más grupos de píxeles en la primera trama realizando las etapas de:
- 35 identificar al menos un grupo de píxeles que representan una región de primer plano de la primera trama y al menos un grupo de píxeles que representan una región de segundo plano de la primera trama; segmentar regiones de primer plano en manchas de primer plano, en el que cada mancha de primer plano representa un objeto representado en la primera trama; y actualizar una imagen de segundo plano de la escena basándose en las regiones de segundo plano identificadas en la primera trama.
- 40
15. El sistema de la reivindicación 14, en el que el motor de visión de ordenador está configurado además para:
- actualizar un mapa con comentarios de la escena representada por el flujo de vídeo usando los resultados de las etapas de generar un modelo de búsqueda que almacena una o más características asociadas con cada objeto identificado;
- 45 clasificar cada uno de los objetos usando un clasificador entrenado; y rastrear, en una segunda trama, cada uno de los objetos identificados en la primera trama usando el modelo de búsqueda, en el que el mapa con comentarios describe una geometría tridimensional de la escena que incluye una posición tridimensional estimada de los objetos identificados y una posición tridimensional estimada de una pluralidad de objetos representados en la imagen de segundo plano de la escena, y en el que la etapa de construir representaciones semánticas comprende además analizar las representaciones semánticas construidas para patrones de conducta reconocibles usando análisis semántico latente.
- 50

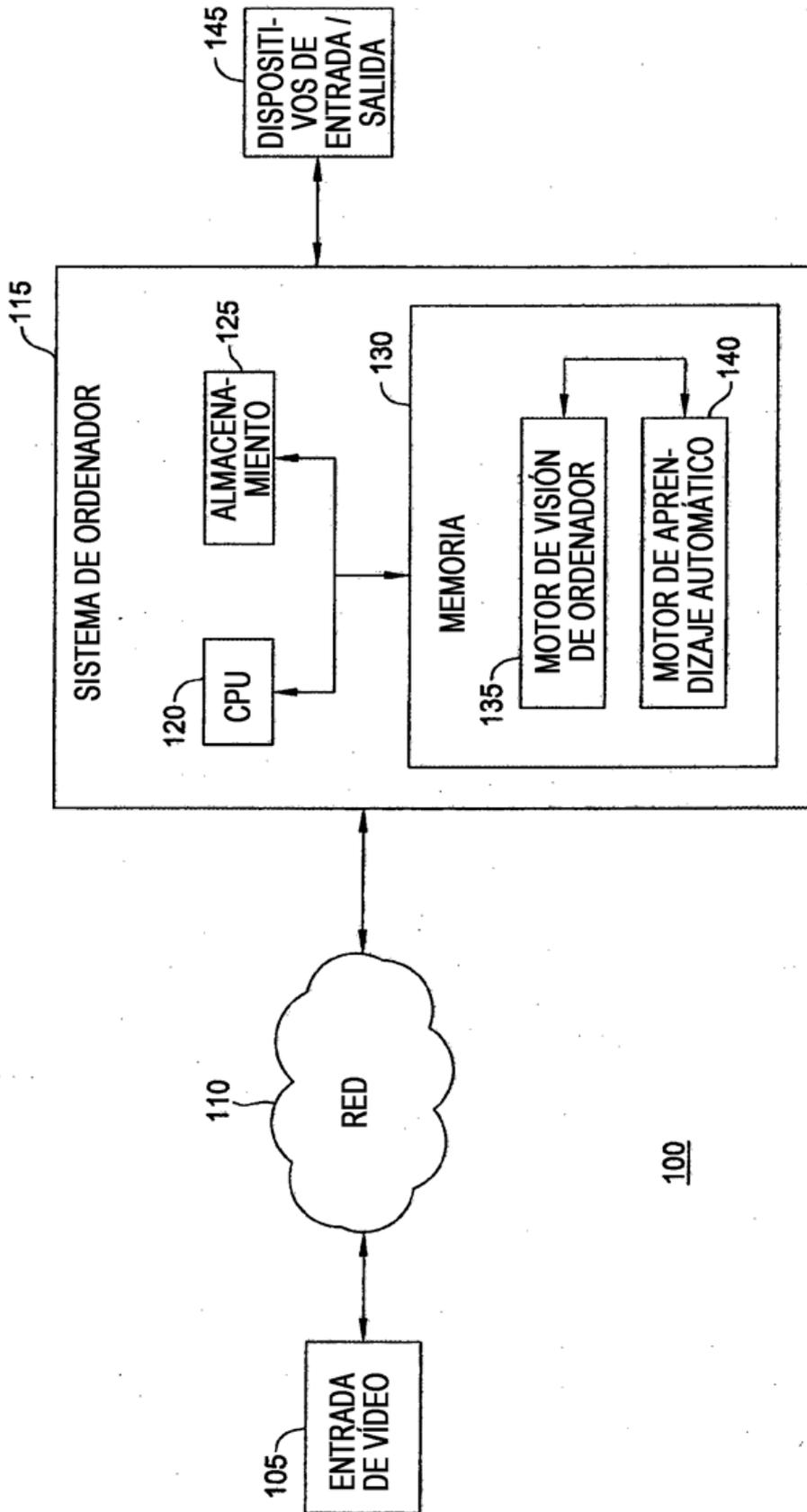


FIG. 1



200

FIG. 2

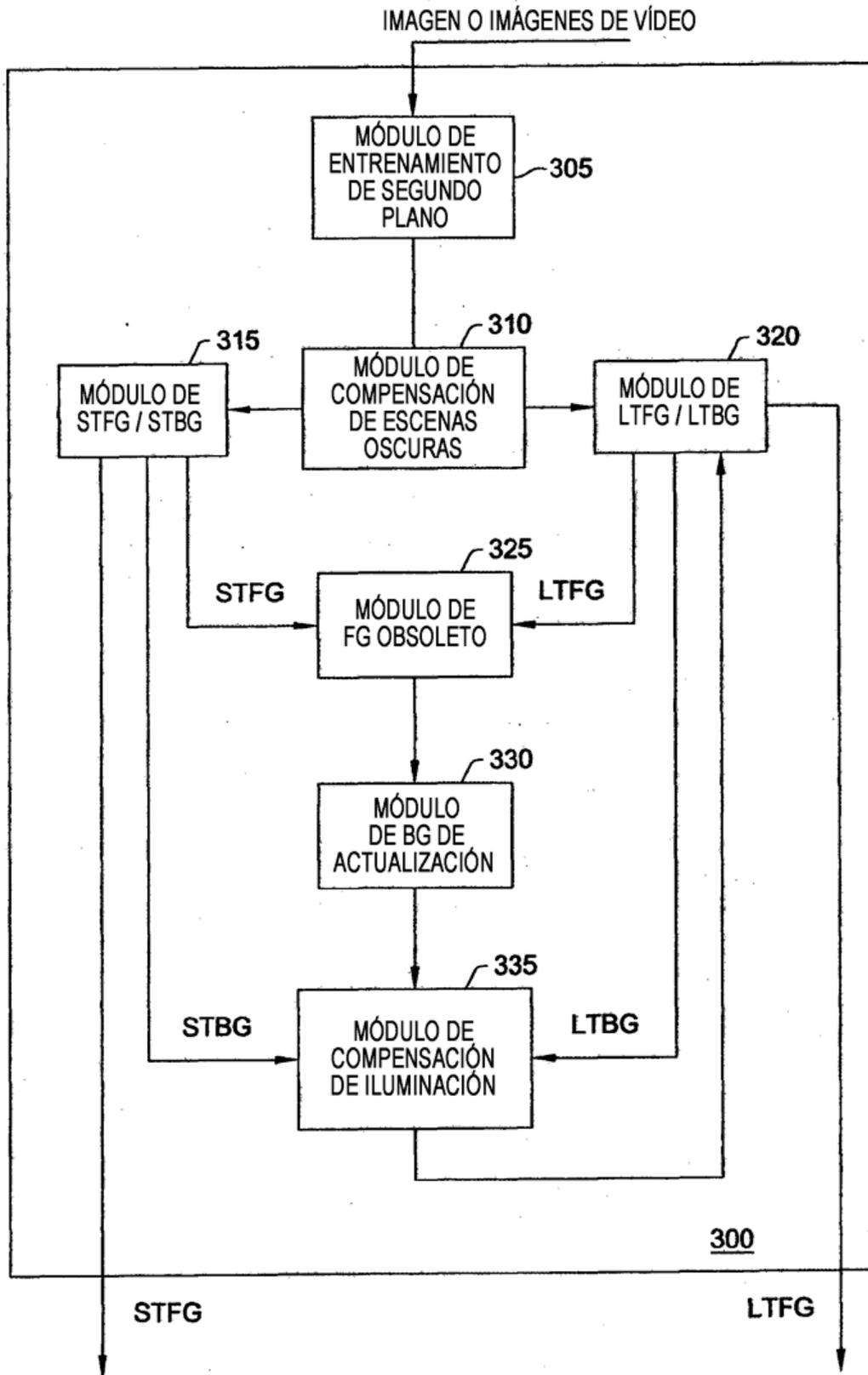


FIG. 3

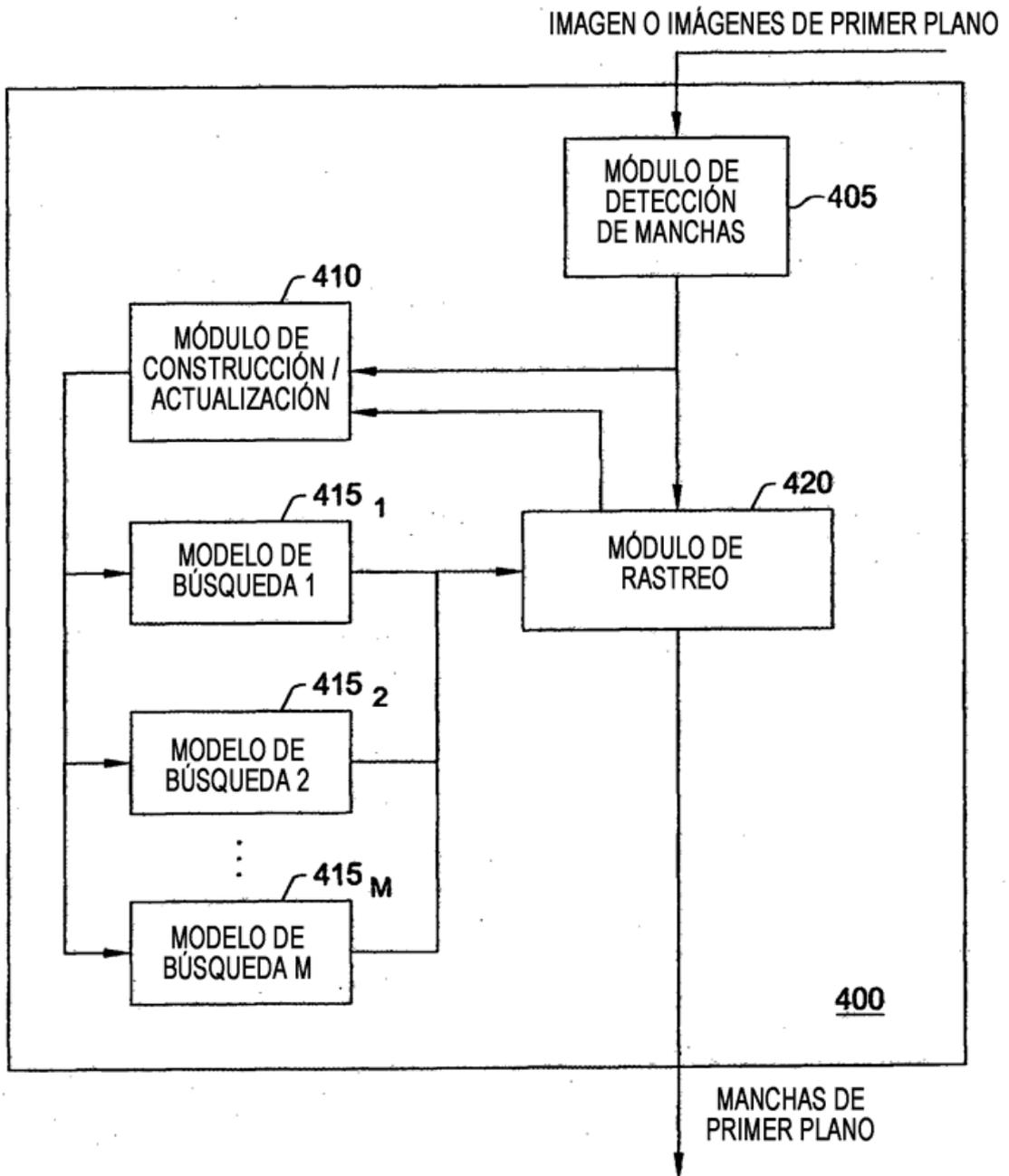


FIG. 4

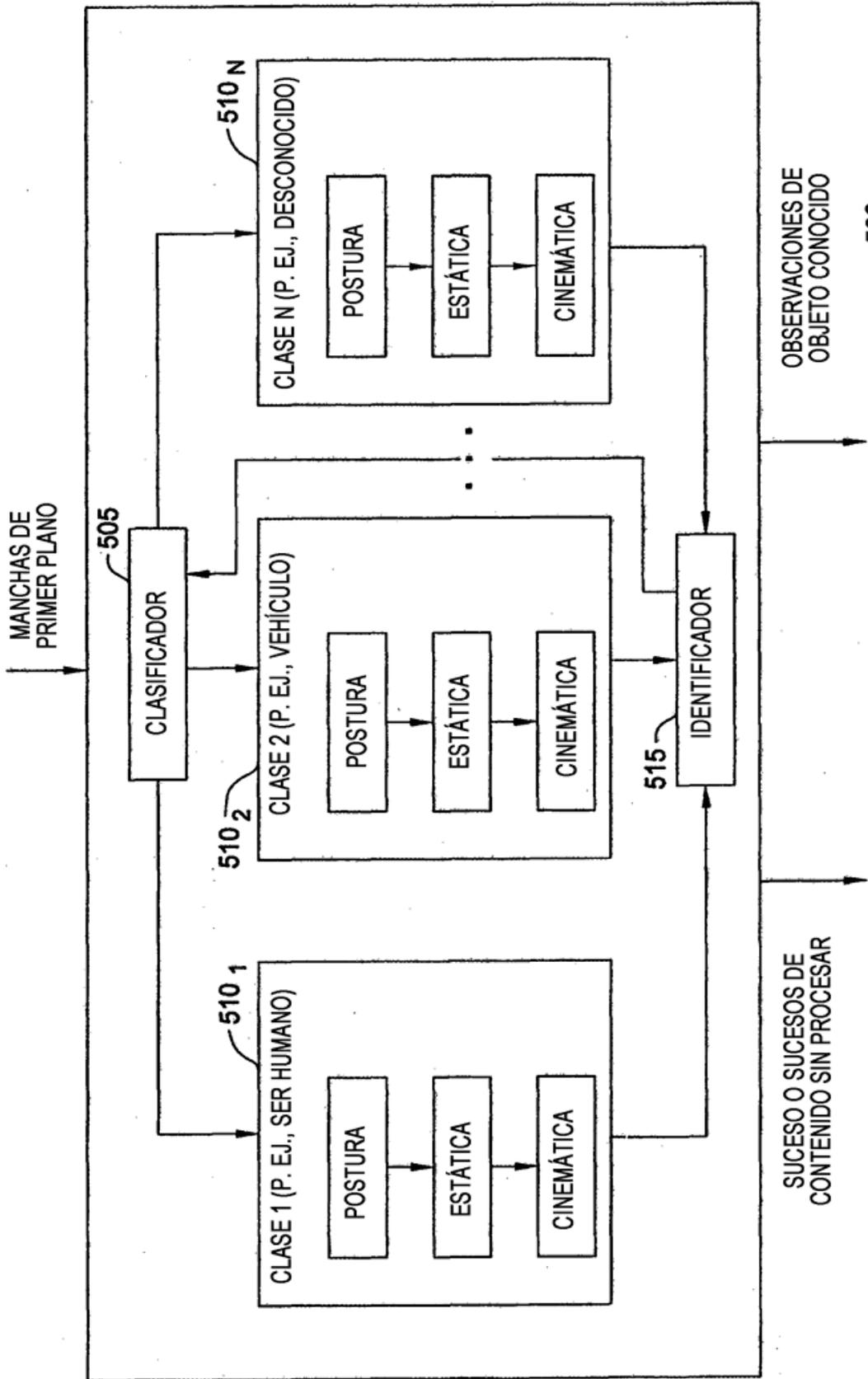


FIG. 5

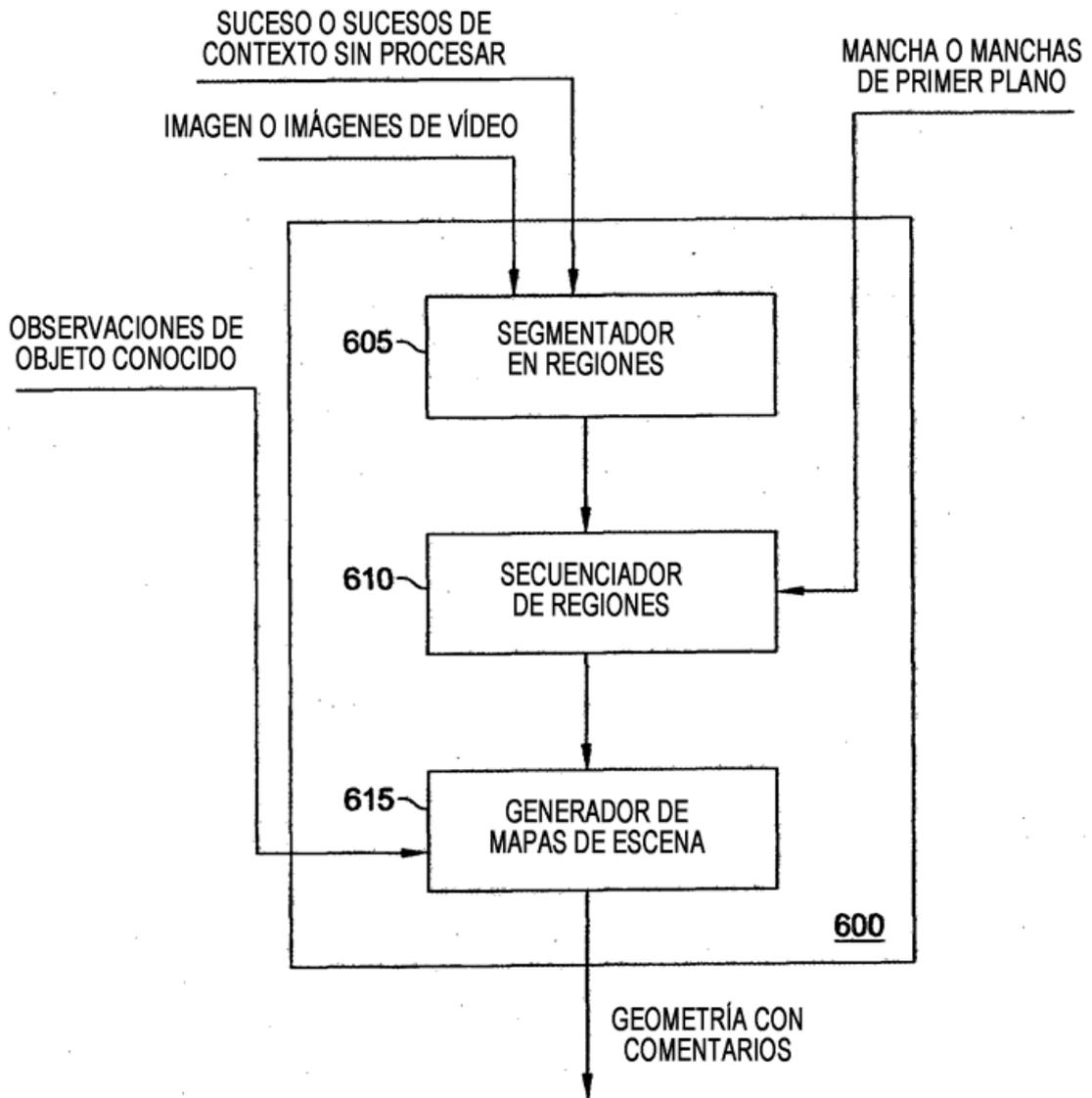


FIG. 6

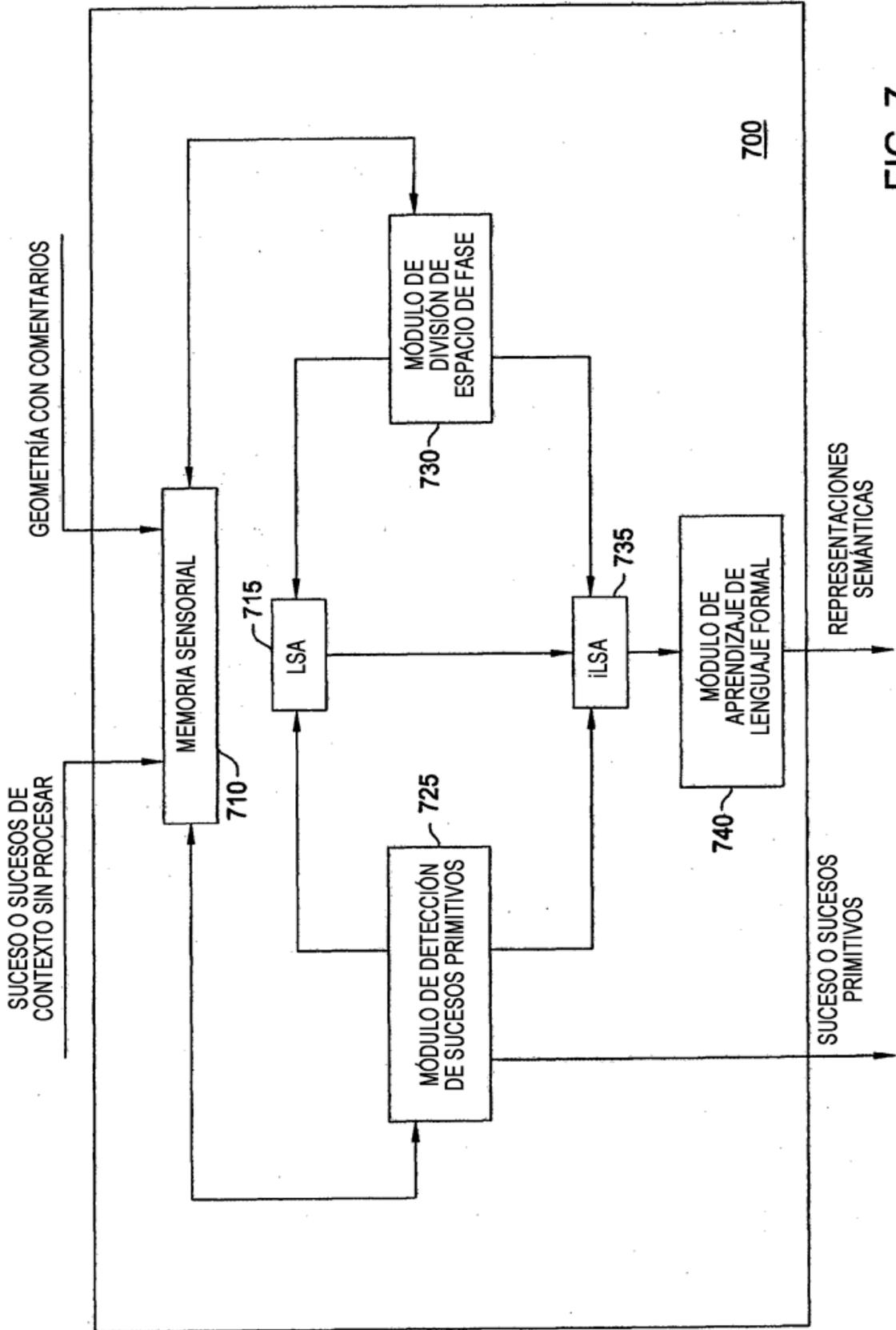


FIG. 7

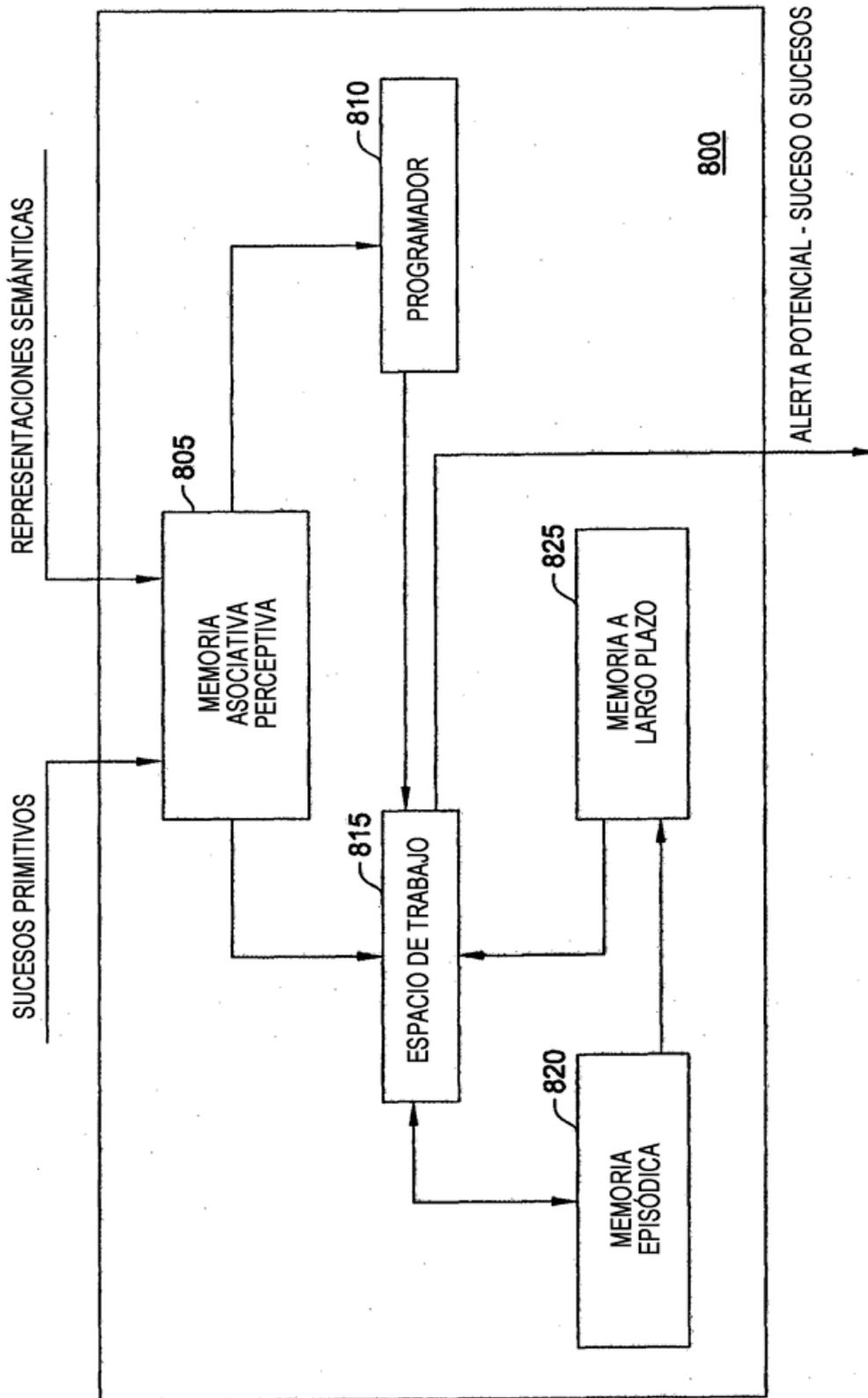


FIG. 8

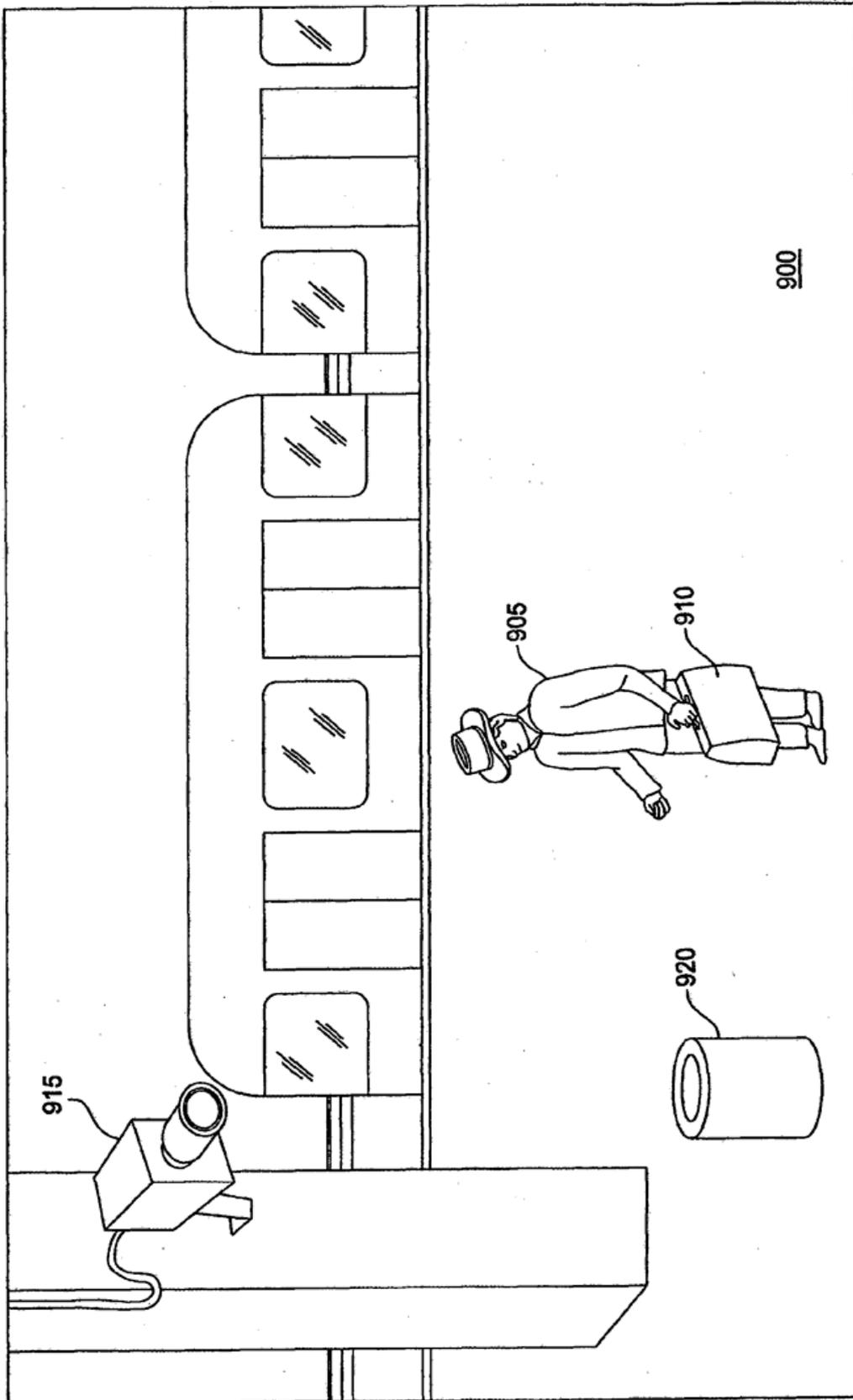


FIG. 9A

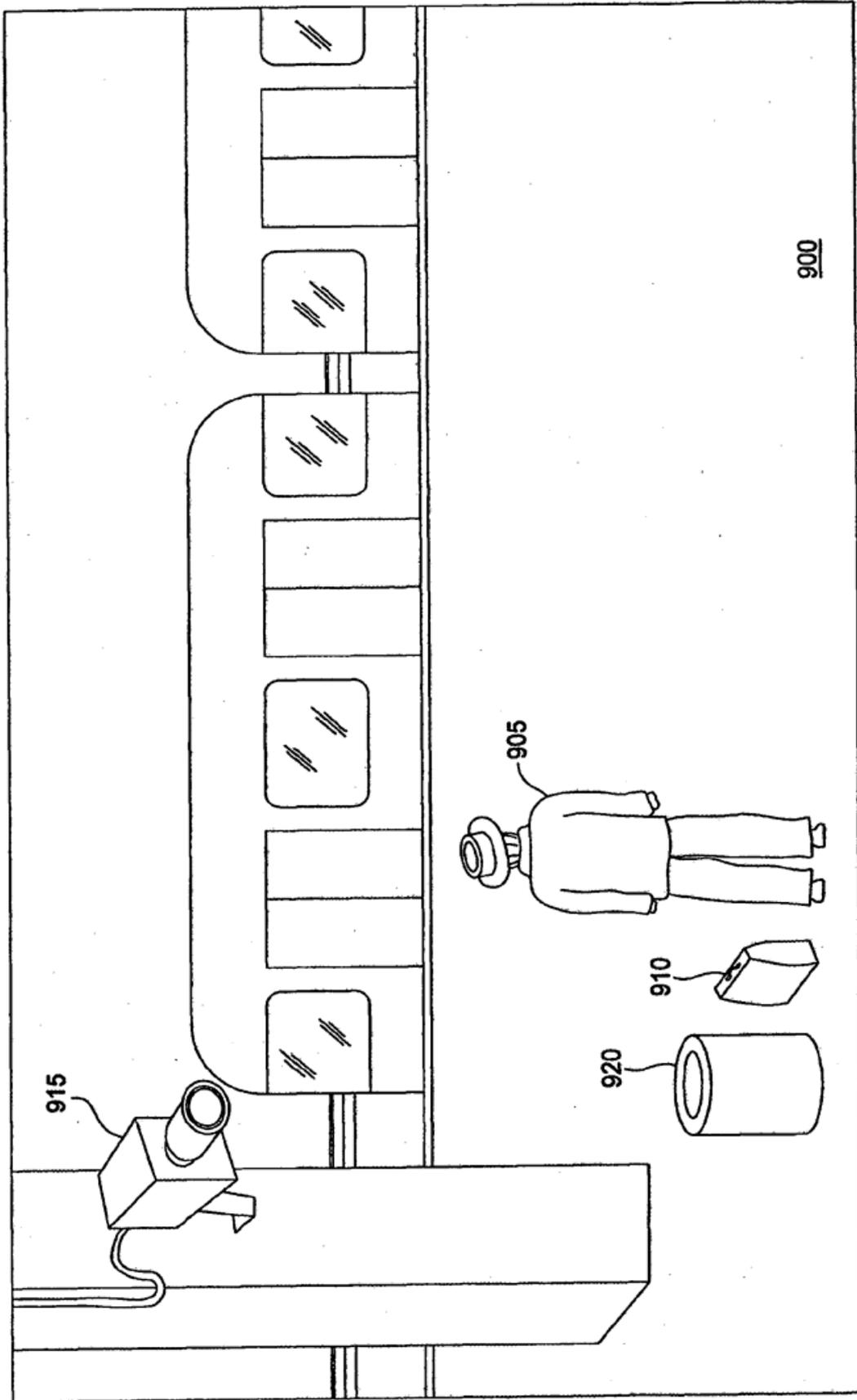


FIG. 9B

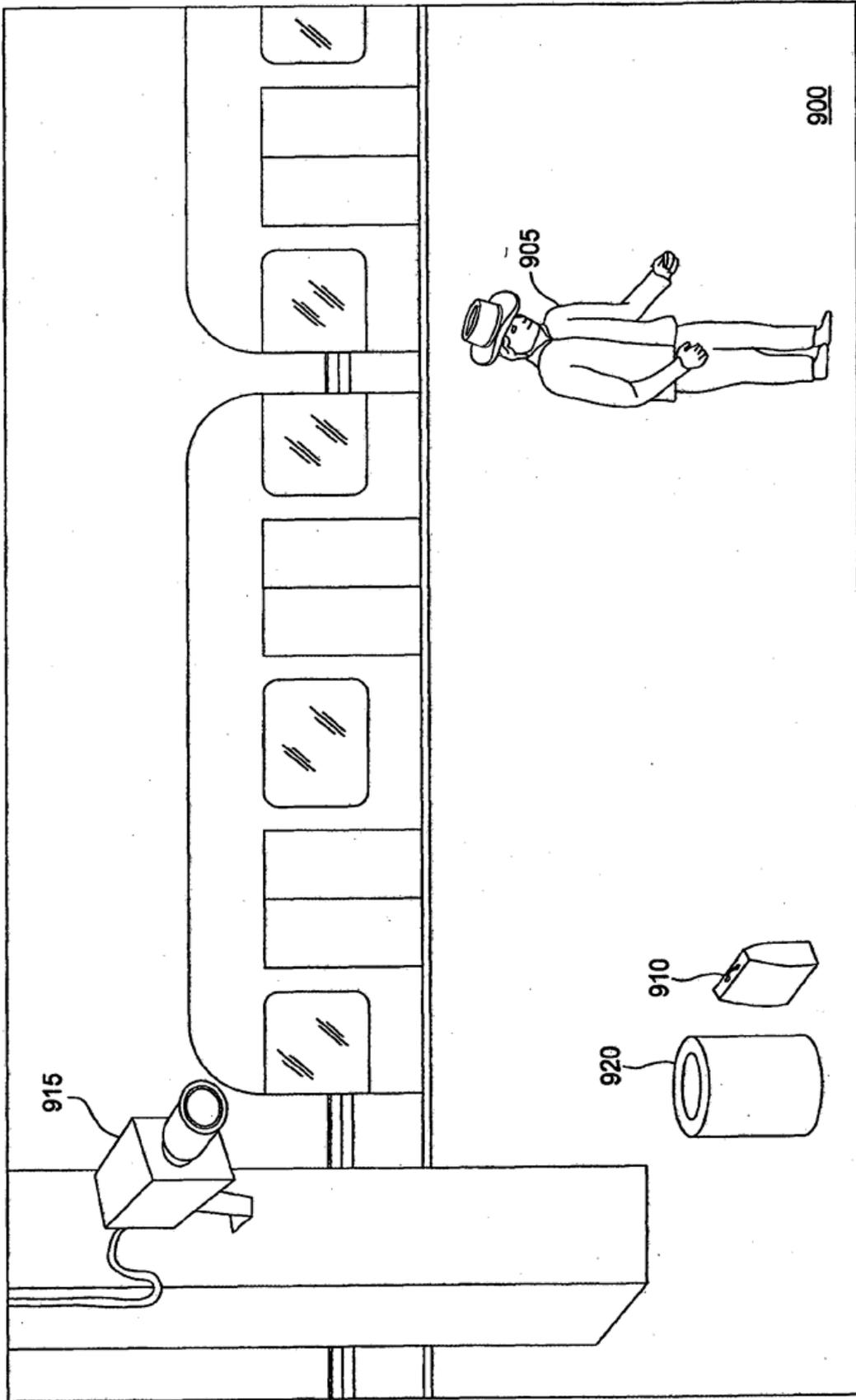


FIG. 9C