

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 767 338**

51 Int. Cl.:

**G01S 5/02** (2010.01)

**G01S 5/20** (2006.01)

**G06T 7/55** (2007.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **07.11.2016 PCT/GB2016/053482**

87 Fecha y número de publicación internacional: **11.05.2017 WO17077348**

96 Fecha de presentación y número de la solicitud europea: **07.11.2016 E 16795405 (6)**

97 Fecha y número de publicación de la concesión europea: **01.01.2020 EP 3371619**

54 Título: **Detección de drones**

30 Prioridad:

**06.11.2015 GB 201519634**

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

**17.06.2020**

73 Titular/es:

**SQUAREHEAD TECHNOLOGY AS (100.0%)  
PO Box 13, Nydalen  
0410 Oslo, NO**

72 Inventor/es:

**HAFIZOVIC, INES;  
NYVOLD, STIG OLUF;  
AASEN, JON PETTER HELGESEN;  
DALENG, JOHANNES ALMING y  
OLSEN, FRODE BERG**

74 Agente/Representante:

**ISERN JARA, Jorge**

**ES 2 767 338 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**

Detección de drones

5 La presente invención hace referencia a la detección de vehículos aéreos no tripulados dentro de una zona de interés utilizando tanto datos de audio como de vídeo.

10 En los últimos años se ha investigado y desarrollado de forma muy abundante sobre los vehículos aéreos no tripulados (UAVs), comúnmente llamado "drones". Dichos drones suelen ser, aunque no siempre lo son, pequeños quadcopters, es decir, helicópteros multirrotor que son levantados y propulsados por la acción de cuatro rotores. Sin embargo, cada vez es más evidente que estos drones representan una amenaza potencial, tanto para la privacidad como para la seguridad públicas.

15 Los drones como los descritos anteriormente se pueden comprar fácilmente en una amplia variedad de tiendas de electrónica comunes y también vía Internet en un proceso con poco o ningún escrutinio de las autoridades. Se ha discutido en gran medida sobre el hecho de que estos drones podrían llegar a ser utilizados para fines nefastos, como, por ejemplo, para transportar material explosivo, biológico o radiactivo. Existen también ciertas preocupaciones relacionadas con la privacidad, pues dichos drones pueden estar equipados con dispositivos de vigilancia, como cámaras y/o micrófonos que podrían utilizarse para espiar a los miembros de establecimientos públicos o privados.

20 Además de ello, dichos drones son, a menudo, muy pequeños y vuelan a altitudes tan bajas que los sistemas de detección de aeronaves convencionales son incapaces de localizarlo. El Solicitante considera que sería muy beneficioso detectar, clasificar y rastrear tales vehículos aéreos no tripulados.

25 El documento GB2103341 describe un lanzador de cohetes en donde el control del cohete se consigue mediante un medio de detección pasiva primaria, como pueda ser un conjunto de micrófonos. El lanzador de cohetes incluye, además, un sistema óptico o térmico utilizado para comparar la imagen objetivo con una serie de objetivos adecuados almacenados.

30 Vista desde un primer aspecto, la presente invención proporciona un sistema para la detección, la clasificación y el rastreo de vehículos aéreos no tripulados en una zona de interés, comprendiendo dicho sistema:

35 al menos una matriz de micrófonos que incluye varios micrófonos, estando la, al menos, una matriz de micrófonos dispuesta para proporcionar datos de audio;

Al menos una cámara dispuesta para proporcionar datos de vídeo; Y

40 al menos un procesador dispuesto para procesar los datos de audio y los datos de video para generar un mapa de probabilidad de detección espacial que comprende un conjunto de celdas espaciales, en donde el procesador asigna una puntuación de probabilidad a cada celda existente dentro del conjunto de celdas espaciales, siendo dicha puntuación de probabilidad una función de:

45 una puntuación de análisis de audios generada por un algoritmo de análisis de audios, comprendiendo dicho algoritmo de análisis de audios la comparación de los datos de audio correspondientes a la celda espacial con una biblioteca de firmas de audio;

Una puntuación de intensidad de audio generada mediante la evaluación de una amplitud de, al menos, una porción de un espectro de los datos de audio correspondientes a la celda espacial; Y

50 una puntuación de análisis de vídeo generada utilizando un algoritmo de procesamiento de imágenes para analizar los datos de vídeo correspondientes a la celda espacial,

55 en donde el sistema está dispuesto para indicar que un vehículo aéreo no tripulado ha sido detectado en una o más celdas espaciales dentro de la zona de interés si la puntuación de probabilidad asignada a dicha o dichas celdas espaciales excede un umbral de detección predeterminado.

60 Por consiguiente, aquellos expertos en la técnica podrán apreciar que la presente invención proporciona un sistema que monitoriza la zona de interés (generalmente, aunque no necesariamente, un área próxima a la localización de dicho sistema) para detectar vehículos aéreos no tripulados o "drones". El sistema está configurado de tal manera que el área monitorizada se divide en una cierta cantidad de celdas espaciales, que son analizadas utilizando los criterios descritos anteriormente para determinar una probabilidad compuesta de que un dron está presente en dicha celda. La indicación de la presencia de un vehículo aéreo no tripulado en una o más de una celda espacial puede constituir una advertencia apropiada, como pueda ser una alerta visual o auditiva, que se proporciona, por ejemplo, a un operario humano o a otros sistemas de seguridad adicionales implementados en ordenador.

65

También será apreciado por aquellos expertos en la técnica que la evaluación de la amplitud de una porción de un espectro de los datos de audio incluye la evaluación de la amplitud en sí misma, pero se debería entender que también incluye la evaluación de la potencia de la porción del espectro o de otra cantidad derivada del mismo.

5 Por consiguiente, al verla desde un segundo aspecto, la presente invención proporciona un sistema para la detección, la clasificación y el rastreo de vehículos aéreos no tripulados en una zona de interés, comprendiendo el sistema:

al menos una matriz de micrófonos que incluye varios micrófonos, estando la, al menos, una matriz de micrófonos dispuesta para proporcionar datos de audio;

10

Al menos una cámara dispuesta para proporcionar datos de vídeo; Y

15

al menos un procesador dispuesto para procesar los datos de audio y los datos de vídeo para generar un mapa de probabilidad de detección espacial que comprende un conjunto de celdas espaciales, en donde el procesador asigna una puntuación de probabilidad a cada celda existente dentro del conjunto de celdas espaciales, siendo dicha puntuación de probabilidad una función de:

20

una puntuación de análisis de audios generada por un algoritmo de análisis de audios, comprendiendo dicho algoritmo de análisis de audios la comparación de los datos de audio correspondientes a la celda espacial con una biblioteca de firmas de audio;

25

Una puntuación de intensidad de audio generada mediante la evaluación de una amplitud de, al menos, una porción de un espectro de los datos de audio correspondientes a la celda espacial; Y

30

una puntuación de análisis de vídeo generada utilizando un algoritmo de procesamiento de imágenes para analizar los datos de vídeo correspondientes a la celda espacial,

en donde el sistema está dispuesto para indicar que un vehículo aéreo no tripulado ha sido detectado en una o más celdas espaciales dentro de la zona de interés si la puntuación de probabilidad asignada a dicha o dichas celdas espaciales excede un umbral de detección predeterminado.

35

Aquellos expertos en la técnica apreciarán que una matriz de micrófonos cuenta con una funcionalidad similar a la de un micrófono omnidireccional, sin embargo, también es capaz de señalar la dirección de una fuente de sonido. Las matrices de micrófonos 2D típicas pueden contar con un rango de posicionamiento de 120° tanto en dirección x- como en dirección y- y con un rango de detección de varios cientos de metros, proporcionando un área de cobertura muy considerable dentro de la zona de interés. Una matriz de micrófonos 3D típica, lo cual puede hacer referencia (pero no se limita a) matrices de micrófonos esféricos, puede tener un rango de posicionamiento verdaderamente omnidireccional (a pesar de que dichas matrices de micrófonos 3D aún poseen un rango de detección limitado). Si bien algunos de los sistemas de acuerdo con la invención pueden detectar drones utilizando una sola matriz de micrófonos, en un conjunto de realizaciones, el sistema comprende varias matrices de micrófonos dispuestas en ubicaciones físicas diferentes, estando cada una de las matrices de micrófonos dispuestas para capturar datos de audio. Ello permite proporcionar diferentes puntos de vista de la zona de interés para potenciar las capacidades del sistema.

40

45

Sin embargo, en un subconjunto de dichas realizaciones, los datos de audio de, al menos, dos matrices de micrófonos adyacentes se combinan con el objetivo de simular una única y más grande matriz de micrófonos. La matriz de micrófonos más grande resultante puede ser más adecuada para la detección de ciertos rangos de frecuencia particularmente bajos. Al combinar matrices de micrófonos de esta forma también puede permitir un procesamiento colaborativo que potencia la resolución y/o el rango en el que pueden ser detectados los drones al combinar las capacidades de las múltiples matrices de micrófonos. Se apreciará que las matrices de micrófonos pueden contar con cualquier geometría y tamaño siempre que sean adecuados y que la invención no se limita a una configuración concreta de la matriz de micrófonos.

50

55

Contar con varias matrices de micrófonos en diferentes ubicaciones físicas también puede ayudar a la hora de determinar cómo de lejos está un vehículo aéreo no tripulado y, en al menos en algunas realizaciones, el sistema está dispuesto para determinar la profundidad en la zona de interés utilizando datos de audio procedentes de varias matrices de micrófonos. En dichas realizaciones, cada matriz de micrófonos proporciona un ángulo de detección relativo a su propia posición y la combinación de estos ángulos de detección puede proporcionar la distancia absoluta hasta un vehículo aéreo no tripulado, por ejemplo, valiéndose de la triangulación, del tiempo de vuelo, de la potencia diferencial recibida, de las diferencias existentes en las envolturas espectrales de varias señales recibidas, del efecto Doppler, etc. En algunas realizaciones potencialmente coincidentes, el sistema está dispuesto para determinar la profundidad dentro de la zona de interés utilizando datos de vídeo. En dicho caso, se puede utilizar el tamaño de un objeto detectado dentro del campo de visión de la cámara o la diferencia existente en el tamaño de un objeto dentro del campo de visión de varias cámaras para determinar la profundidad dentro de la zona de interés.

60

65

Si bien el sistema puede detectar drones utilizando una sola cámara, en un conjunto de realizaciones, el sistema comprende varias cámaras dispuestas en ubicaciones físicas diferentes. Tal y como ocurre con la/las matriz/matrices de micrófonos, contar con varias cámaras proporciona puntos de vista adicionales de la zona de interés.

5 Si bien las múltiples cámaras podrían utilizarse para producir datos de vídeo correspondientes a puntos de vista de la zona de interés completamente separados y no coincidentes, en una serie de realizaciones, los datos de vídeo de, al menos, dos de las cámaras se registran conjuntamente para generar un mapeo de la imagen entre ellos. En dichas realizaciones, hay al menos un cierto grado de superposición entre los puntos de vista cubiertos por cada cámara y, por consiguiente, es posible utilizar un algoritmo de corrección, conocido en el contexto de la técnica por su función para crear mapeos de imágenes. Cada matriz de micrófonos puede estar asociada con una o más cámaras de manera que el "plano de sonido" de la matriz de micrófonos esté combinado con la(s) cámara(s) asociada(s) con el mismo. La calibración y el corrección automático entre cámaras puede llevarse a cabo utilizando algoritmos de procesamiento de imágenes, conocidos en la técnica por sí mismos y que pueden ser utilizados para registrar o "aunar" planos de sonido de diferentes matrices de micrófonos. Es esta una cuestión ventajosa, pues permite, por ejemplo, que el sistema esté al tanto por adelantado de un objeto detectado en los datos de vídeo procedentes de una primera cámara que viaja hacia el punto de vista de una segunda cámara.

20 Si bien una sola cámara tan solo proporciona una proyección en 2D del espacio en 3D, en una serie de realizaciones, varias cámaras forman una disposición estereoscópica que detecta la profundidad dentro de la zona de interés. En un subconjunto de dichas realizaciones, las varias cámaras forman varias disposiciones estereoscópicas dispuestas para detectar diferentes profundidades dentro de la zona de interés. Dicha disposición ventajosa proporciona información de la profundidad desde varias vistas diferentes, lo cual mejora así la certeza del valor de profundidad determinado por el procesador.

25 De forma adicional o alternativa, en al menos alguna de las realizaciones, se utilizan los datos de audio procedentes de, al menos, una matriz de micrófonos para mejorar la detección de profundidad llevada a cabo utilizando varias cámaras. Ello puede lograrse, por ejemplo, midiendo la diferencia en el tiempo existente entre la detección de un dron por parte de una cámara y la detección por parte de una matriz de micrófonos. Dado que se conocen la velocidad del sonido y la distancia existente entre la cámara y la matriz de micrófonos, es posible determinar de forma precisa la distancia existente hasta el dron partiendo de la diferencia en el tiempo que se produce entre el momento en el que la cámara "ve" un evento (como la aparición de un dron o su movimiento hacia una cierta dirección) y el momento en el que la matriz de micrófonos "oye" ese mismo evento.

35 Si bien la(s) cámara(s) utilizada(s) por el sistema pueden ser unidades independientes y separadas de la/las matriz/matrices de micrófonos, denominadas en lo sucesivo "cámaras externas", en un conjunto de realizaciones, al menos una matriz de micrófonos incluye una cámara. Dicha cámara "integrada" puede estar ubicada en la superficie de la matriz de micrófonos, por ejemplo, en el centro de la matriz de micrófonos, rodeada por micrófonos individuales o en alguna otra ubicación fija y conocida en relación con la matriz de micrófonos. Esto crea un mapeo relativamente directo entre dicha cámara y la matriz de micrófonos asociada. En algunas realizaciones, cada matriz de micrófonos incluye una cámara.

Ello, por supuesto, no excluye la existencia de más cámaras externas que puedan estar ubicadas en algún otro lugar con más puntos de vista de la zona de interés.

45 Dado que la/las matriz/matrices de micrófonos y la(s) cámara(s) son normalmente posicionadas de forma estratégica por el usuario y que generalmente son estáticas, la relación espacial entre ambos elementos es conocida. En consecuencia, en una serie de realizaciones, al menos dos matrices de micrófonos y/o cámaras son mapeadas entre sí utilizando una relación espacial conocida entre las ubicaciones físicas de la/las matriz/matrices de micrófonos y/o la(s) cámara(s), de manera que dicha(s) matriz/matrices de micrófonos y/o cámara(s) comparten un sistema de coordenadas común. Dichas coordenadas pueden ser coordenadas del sistema de coordenadas universal transversal de Mercator (UTM) tal y como se explicará con detalle a continuación.

55 En algunas realizaciones, el sistema comprende un subsistema de sensor periférico, en donde el subsistema de sensor periférico comprende, al menos, un elemento de un grupo que comprende: un sensor de navegación por satélite global; Un giroscopio; Un magnetómetro; Un acelerómetro; Un reloj; Un anemómetro electrónico; Y un termómetro. En algunas de dichas realizaciones, el subsistema sensor periférico está integrado en una o en más de una matriz de micrófonos. Un sistema proporcionado con el mencionado subsistema sensor periférico puede llegar a utilizar los componentes que contiene para mejorar las capacidades de detección del sistema. Por ejemplo, el sistema puede ser dispuesto para utilizar datos procedentes del subsistema sensor periférico para compensar las variaciones existentes en los parámetros de propagación del sonido, como la velocidad y la temperatura del viento.

60 El número y la densidad de las celdas espaciales pueden ser predeterminados e inamovibles. En una serie de realizaciones, el conjunto de celdas se genera automáticamente. Dicha generación puede estar basada en factores como la resolución, la relación señal ruido (SNR) de la salida, la ganancia de la matriz de micrófonos, la capacidad del procesador, etc. En otro conjunto de realizaciones, la generación del conjunto de celdas espaciales puede ser iniciada por el usuario, por ejemplo, dividiendo de forma manual la zona de interés en células espaciales individuales. En un

conjunto de realizaciones, el procesador está dispuesto de forma selectiva para aumentar un número de celdas espaciales en, al menos, un subconjunto de dicha zona de interés si la puntuación de probabilidad asignada a una o más de una celda espacial en dicho subconjunto llegara a exceder un umbral de cambio de densidad de celda predeterminado. En un subconjunto de realizaciones, el umbral de cambio de densidad de celda es inferior al umbral de detección. Ello, ventajosamente, permite que se lleve a cabo un escaneado preliminar más disperso y "rudimentario" y, entonces, si el sistema determina que un dron podría estar presente, pero está inseguro debido a que el número de celdas es insuficiente, podría llevarse a cabo al menos una operación de generación de haces en donde los datos de varios micrófonos de una o más matrices son filtrados de acuerdo con un algoritmo de generación de haces. En ciertas disposiciones, un algoritmo disperso puede elegir inicialmente procesar solo un subconjunto de micrófonos, un subconjunto de matrices de micrófonos o un subconjunto de celdas espaciales en las que se divide la zona de interés. El aumento de la densidad solo en la región en la que se supone que hay un dron puede facilitar un ahorro significativo en la potencia de procesamiento necesaria para analizar las celdas espaciales en las que no se ha detectado ningún dron.

En un conjunto de realizaciones, el procesador está dispuesto de forma selectiva para perfeccionar la resolución de al menos una matriz de micrófonos y/o cámara si la puntuación de probabilidad asignada a dicha una o más celdas espaciales excede un umbral de cambio de resolución predeterminado. En un subconjunto de realizaciones, el umbral de cambio de resolución es inferior al umbral de detección. Ello, ventajosamente, permite que se lleve a cabo un escaneado preliminar más disperso y "rudimentario" y, entonces, si el sistema determina que un dron podría estar presente, pero está inseguro debido a que la resolución de una matriz de micrófonos, de una cámara o ambas opciones, puede aplicar un aumento de la resolución. El perfeccionamiento de la resolución puede lograrse llevando a cabo un procedimiento de procesamiento adicional. Por ejemplo, los datos procedentes de la matriz de micrófonos pueden ser analizados inicialmente para detectar la presencia de un dron de manera que los haces individuales analizados sean primero separados en ángulos de 10°, pero si el sistema determina que existe una posibilidad significativa de que haya un dron presente (es decir, si la puntuación de probabilidad excede el umbral de cambio de resolución), la resolución puede ser incrementada de manera que los haces estén separados por ángulos de, por ejemplo, 1°. En dicho caso se podría aplicar un algoritmo de generación de haces computacionalmente más avanzado o complejo (por ejemplo, un "generador de haces superdireccional" con una mayor resolución espacial para lograr un posicionamiento más preciso de un dron. De forma similar, los datos de vídeo de la cámara pueden no procesarse píxel por píxel, sino, por ejemplo, pueden ser submuestreados de manera que el procesador tan solo tenga que analizar, por ejemplo, uno de cada dos píxeles o una de cada dos líneas de escaneo, etc. para reducir los requisitos de potencia de procesamiento promedios. A continuación, si se considera que se ha podido detectar un dron, el procesador puede comenzar a muestrear cada píxel disponible o, al menos, cada píxel existente en las cercanías de la ubicación estimada del dron.

De forma similar, en un conjunto de realizaciones, al menos una cámara está dispuesta para hacer zoom en un área dentro de la zona de interés si la puntuación de probabilidad asignada a dicha o a dichas celdas espaciales excede un umbral de zoom predeterminado. En un subconjunto de dichas realizaciones, el umbral de cambio de zoom es inferior al umbral de detección. En dicha situación, el sistema puede cambiar el zoom de una cámara para inspeccionar en mayor medida un área dentro de la zona de interés si se considera que podría haber presencia en ella de un dron. Dicho zoom puede ser un zoom óptico en donde las ópticas de los objetivos se mueven para cambiar la distancia focal de la cámara o también puede ser un zoom digital en donde el zoom se lleva a cabo artificialmente ampliando la imagen. Si bien un zoom digital puede reducir la resolución, en ciertas circunstancias puede resultar útil. Sin embargo, es preferible proporcionar la(s) cámara(s) con funcionalidad de zoom óptico.

En un conjunto de realizaciones, el conjunto de celdas espaciales está mapeado adicionalmente a datos de calibración que comprenden varias coordenadas del sistema de posicionamiento global. El solicitante ha apreciado que dicho sistema de coordenadas común permite la traducción a coordenadas del sistema de coordenadas universal transversal de Mercator (UTM) correctas, una proyección cartográfica ampliamente utilizada para traducir la longitud y la latitud de una posición en la Tierra esférica a una posición en un plano representado en 2D, como un mapa. Por supuesto, existen otras proyecciones cartográficas y los expertos en la materia apreciarán que otras transformaciones apropiadas serán igualmente factibles.

En un conjunto de realizaciones, los datos de calibración se generan previamente en un dispositivo GNSS (sistema global de navegación por satélite) de prueba ubicado en un dron de calibración, estando dicho dron de calibración dispuesto para atravesar el conjunto de celdas espaciales. El dron de calibración puede producir una firma de audio conocida que el sistema puede detectar. De forma adicional, o alternativa, el dron de calibración puede tener una firma visual conocida, como una fuente de luz particularmente brillante, adjunta al mismo y que puede ser detectada por el sistema. Al preparar al dron de calibración para que pueda volar a través de la zona de interés de manera que atraviese parte o la totalidad del conjunto de celdas espaciales y al comparar un registro de las coordenadas GNSS registrado por la zona de calibración con las celdas espaciales en las que la zona de calibración es detectada por el sistema, es posible crear una relación uno a uno de las celdas espaciales y las coordenadas GNSS apropiadas. Ello requiere la sincronización entre los datos del dispositivo GNSS y el sistema de detección. Se apreciará que los términos "sistema de navegación por satélite global" y "GNSS", tal y como se usan en el presente documento, no se limitan a ningún sistema de posicionamiento concreto y debe entenderse que incluyen todos los sistemas de posicionamiento adecuados, como puedan ser el Global Positioning System (GPS), el Global Navigation Satellite System (GLONASS,

Galileo o BeiDou. La técnica de navegación por satélite Real Time Kinematic (RTK), conocida en la técnica por sí misma, también puede utilizarse para aumentar la precisión de los datos de posicionamiento obtenidos por el sistema.

5 Si bien es posible alinear los datos procedentes de la/las matriz/matrices de micrófonos y de la(s) cámara(s) de forma temporal durante el postprocesamiento, en un conjunto de realizaciones, cada una de las matrices de micrófonos y/o cámaras están sincronizadas en el tiempo. En un subconjunto de dichas realizaciones, la sincronización en el tiempo se consigue enviando a cada matriz de micrófonos y/o cámara una marca de tiempo generada por un servidor central. Al sincronizar las matrices de micrófonos y/o las cámaras, es posible analizar los datos procedentes de cada fuente en tiempo real, por lo que se puede saber con certeza que los datos de cada fuente corresponden a los datos de cada una de las otras fuentes sincronizadas. La marca de tiempo generada por el servidor central puede ser enviada una vez durante la inicialización del sistema o también puede ser enviada de forma continua durante el funcionamiento del sistema, tanto periódicamente como intermitentemente, por ejemplo, según sea preciso. La marca de tiempo puede ser proporcionada de forma alternativa por una fuente no centralizada, como pueda ser la información temporal de GPS o desde una red de datos celulares.

15 En un conjunto de realizaciones, los datos de audio procedentes de al menos una matriz de micrófonos se utilizan para guiar el análisis de los datos de vídeo de al menos una cámara. Por consiguiente, si los datos de audio indican que podría darse la presencia de un dron en una región particular de la zona de interés, el procesador puede recibir instrucciones para llevar a cabo un análisis más intensivo de los datos de vídeo correspondientes, por ejemplo, poniendo en práctica algoritmos de procesamiento de imágenes más avanzados en una selección de los píxeles para mejorar el análisis de vídeo.

20 De forma adicional o alternativa, en un conjunto de realizaciones, los datos de vídeo procedentes de al menos una cámara se utilizan para guiar el análisis de los datos de audio de al menos una matriz de micrófonos. De forma similar a la situación descrita anteriormente, si los datos de vídeo indican la posible presencia de un dron en una región particular dentro de la zona de interés, el procesador puede recibir instrucciones para llevar a cabo procedimientos de generación de haces más precisos o de combinación de firmas de audio refinadas adicionales en los datos de audio correspondientes.

25 Los expertos en la técnica apreciarán que existen numerosas técnicas de procesamiento de imágenes y de visión artificial dispuestas para ser utilizadas en las realizaciones de la presente invención. En un conjunto particular de realizaciones, el algoritmo de procesamiento de imágenes comprende:

35 el cálculo de un marco promedio a partir de un subconjunto de cuadros de datos de vídeo recibidos previamente;

La resta de dicho marco promedio de los cuadros de los datos de vídeo recibidos posteriormente para generar una imagen de diferencia; Y

40 la comparación de dicha imagen de diferencia con un umbral dentro de cada celda espacial visual para generar una puntuación para el análisis de vídeo. Por supuesto, se apreciará que la presente invención no está limitada a este algoritmo en particular y que otros algoritmos también pueden utilizarse de forma sencilla permaneciendo dentro del alcance de la invención.

45 En un conjunto de realizaciones, la biblioteca de firmas de audio comprende varias firmas de audio asociadas con los vehículos aéreos no tripulados en varios escenarios. Dichos escenarios pueden, por ejemplo, incluir los sonidos del dron durante el vuelo, el despegue, el aterrizaje, el desplazamiento lateral, el desplazamiento hacia y desde una matriz de micrófonos, en interior, en exterior, etc.

50 Se apreciará que hay una enorme cantidad de modelos diferentes de vehículos aéreos no tripulados o drones disponibles en el mercado y los sonidos producidos por cada modelo pueden variar drásticamente. Por lo tanto, en algunas realizaciones, el algoritmo de análisis de audio comprende la clasificación del vehículo aéreo no tripulado detectado. La clasificación de un vehículo aéreo no tripulado puede, al menos en algunas realizaciones, estar basada en la coincidencia más cercana con respecto a una firma de audio correspondiente a un modelo específico. Esto puede ser particularmente útil a la hora de identificar y rastrear categorías particulares de drones. La clasificación del vehículo aéreo no tripulado puede, de forma adicional o alternativa, llevarse a cabo visualmente y, por consiguiente, en un conjunto de realizaciones potencialmente superpuestas, el algoritmo de procesamiento de imagen comprende la clasificación del vehículo aéreo no tripulado detectado.

60 Si bien se apreciará que en la presente invención podrían aplicarse varias técnicas de análisis diferentes conocidas en la técnica por sí mismas, en al menos algunas realizaciones, el algoritmo de análisis de audio comprende un algoritmo de aprendizaje automático.

65 Ello permite al sistema analizar los datos de audio utilizando el reconocimiento de patrones y modelos estadísticos para generar la puntuación del análisis de audio.

5 En un conjunto de realizaciones, el algoritmo de análisis de audio comprende además compensar una fuente de ruido predeterminada próxima a la zona de interés. Tales fuentes de ruido pueden, a modo de ejemplo, incluir vientos fuertes, ruido producido por el tráfico, sonido de agua fluyendo, etc. Esto permite que el sistema ignore o cancele estas fuentes de sonido para potenciar la SNR del sistema, mejorando así su capacidad de detección. En algunas de dichas realizaciones, el algoritmo de análisis de audio comprende compensar la fuente de ruido predeterminada de forma automática. En dichas realizaciones, el sistema puede autocalibrarse, necesitándose poca o ninguna aportación del usuario para compensar las fuentes de ruido externas no deseadas.

10 En un conjunto de realizaciones, el algoritmo de análisis de audio comprende un algoritmo de gradiente, en donde el algoritmo de gradiente está dispuesto para medir un cambio relativo en una distribución audio-espacial a través de una o más de una celda espacial. En dichas realizaciones, el cambio relativo de la distribución audio-espacial (es decir, los datos de audio a través de la celda o las celdas espaciales) puede ser indicativo de la presencia de un dron.

15 En un conjunto de realizaciones, el procesador está dispuesto para procesar dichos datos de audio y visuales en una serie de periodos de tiempo repetitivos, de manera que procesa datos para cada celda espacial dentro de cada período de tiempo. Se apreciará que esta cuestión proporciona al sistema al menos un modo de operación pseudo-paralela en el que se analiza la totalidad del conjunto de celdas espaciales en cada período de tiempo. Sin embargo, en un conjunto de realizaciones, el procesador está dispuesto para analizar cada celda espacial en paralelo. En dichas realizaciones en las que el procesador es adecuadamente potente, todas las celdas pueden analizarse en paralelo verdaderamente.

20 Los expertos en la técnica apreciarán que la puntuación de probabilidad puede contar con una relación matemática con la puntuación del análisis de audio, la puntuación de intensidad de audio y la puntuación del análisis de vídeo, según corresponda. En algunas realizaciones, la puntuación de probabilidad es un total de la puntuación del análisis de audio, la puntuación de intensidad de audio y la puntuación del análisis de vídeo. De forma alternativa, al menos en algunas realizaciones, la puntuación de probabilidad es una media de la puntuación del análisis de audio, de la puntuación de intensidad de audio y de la puntuación del análisis de vídeo. En algunas de dichas realizaciones, la puntuación de probabilidad es un promedio ponderado de la puntuación del análisis de audio, de la puntuación de la intensidad de audio y de la puntuación del análisis de vídeo. En algunas realizaciones, la función de la puntuación de probabilidad varía dinámicamente durante un funcionamiento regular del sistema.

A continuación, se describirán ciertas realizaciones de la invención, a modo de ejemplo únicamente, con referencia a los dibujos adjuntos, en los que:

35 La figura 1 muestra un vehículo aéreo no tripulado típico a ser detectado por las realizaciones descritas de la presente invención;

40 La figura 2 muestra un sistema de detección de vehículos aéreos no tripulados de acuerdo con una realización de la presente invención;

La figura 3 muestra un conjunto de celdas espaciales utilizado por el procesador del sistema de detección de la figura 2;

45 La figura 4 muestra el vehículo aéreo no tripulado de la figura 1 entrando en la zona de interés del sistema de detección de la figura 2;

La figura 5 muestra un conjunto de celdas espaciales utilizadas por el procesador del sistema de detección de la figura 2 cuando entra el vehículo aéreo no tripulado;

50 La figura 6 muestra el mapa de probabilidad de detección espacial después de que el procesador haya efectuado el análisis;

La figura 7 muestra un ejemplo de proceso de análisis de audio utilizando una biblioteca de firmas de audio;

55 La figura 8 muestra el conjunto de celdas espaciales de la figura 5 habiendo sido refinadas tras la detección del vehículo aéreo no tripulado;

60 La figura 9 muestra un sistema de detección de vehículos aéreos no tripulados de acuerdo con una realización adicional de la presente invención que utiliza múltiples matrices de micrófonos;

La figura 10 muestra un sistema de detección de vehículos aéreos no tripulados de acuerdo con una realización adicional de la presente invención que utiliza múltiples cámaras;

65 La figura 11 muestra los puntos de vista de las cámaras de la figura 10;

La figura 12 muestra el corregistro de los puntos de vista de la figura 11;

La figura 13 muestra la operación de un dron de calibración utilizado para mapear las celdas espaciales a coordenadas GPS del mundo real;

5 La figura 14 muestra cómo las celdas espaciales utilizadas por el procesador de la figura 13 son calibradas utilizando el dron de calibración;

La figura 15 muestra una fuente de ruido constante que puede ser compensada de acuerdo con las realizaciones de la presente invención;

10 La figura 16 muestra un subconjunto de celdas espaciales de la figura 8 refinadas en mayor medida en las proximidades del vehículo aéreo no tripulado detectado; Y

15 La figura 17 muestra un diagrama de bloques de un ejemplo adicional de un proceso de análisis de audio que utiliza un algoritmo de detección y clasificación de características.

La figura 1 muestra un vehículo aéreo no tripulado 50 típico. Este vehículo aéreo no tripulado (VANT) 50 en particular cuenta con un factor de forma tipo quadcopter convencional, en donde el cuerpo del VANT 50 está rodeado por cuatro rotores 52A, 52B, 52C, 52D.

20 Estos VANTs generalmente utilizan un giroscopio para lograr estabilidad, utilizando los datos procedentes del giroscopio para compensar cualquier movimiento lateral no deseado. Un VANT de tipo quadcopter utiliza los rotores 52A, 52B, 52C, 52D en dos pares. Un primer par, que comprende los rotores 52A, 52D, rota en sentido horario, mientras que el segundo par, que comprende los rotores 52B, 52C, rota en sentido antihorario. Cada rotor 52A, 52B, 52C, 52D puede ser controlado de forma independiente para controlar el vuelo del VANT 50. Variar las velocidades de cada uno de los rotores 52A, 52B, 52C, 52D permite la generación de empuje y torque según sea necesario para una ruta de vuelo dada.

30 Dicho tipo de VANT 50 posee una firma de audio (o un conjunto de firmas de audio) que es característica del mismo. Por ejemplo, el sonido de los rotores 52A, 52B, 52C, 52D durante el vuelo contendrá picos a frecuencias específicas dentro del espectro de frecuencias. Tales picos pueden variar en maniobras de vuelo concretas, como: el ajuste de la altitud (al aumentar/disminuir las velocidades de rotación de los rotores 52A, 52B, 52C, 52D por igual); El ajuste del "ángulo" o de la "inclinación" (aumentando la velocidad de rotación de un rotor y disminuyendo la velocidad de rotación de su rotor diametralmente opuesto); O el ajuste de la desviación (aumentando la velocidad de rotación de los rotores que rotan en una dirección y disminuyendo la velocidad de rotación de los rotores que rotan en la dirección opuesta).

35 Cada modelo y diseño de dichos vehículos aéreos no tripulados tendrá diferentes firmas de audio y, por lo tanto, podrán ser identificados como se discutirá más adelante.

40 La figura 2 muestra un sistema de detección de vehículos aéreos no tripulados 2 de acuerdo con una realización de la presente invención. En aras de la claridad, este sistema 2 tan solo cuenta con una matriz de micrófonos 4 y una sola cámara externa 8. La matriz de micrófonos 4 y la cámara 8 están conectadas a un procesador 10, que, en este ejemplo, es un terminal de ordenador.

45 La matriz de micrófonos 4 también cuenta con una cámara integrada 6. Dicha cámara integrada 6 está posicionada en el centro de la matriz de micrófonos 4 y proporciona datos de vídeo que corresponden al mismo punto de vista que los datos de audio proporcionados por la matriz de micrófonos 4. Sin embargo, se apreciará que la cámara integrada 6 no tiene que colocarse, necesariamente, en el centro de la matriz de micrófonos 4 y podría estar, en su lugar, colocada en cualquier otro punto fijo de la matriz de micrófonos 4 o en su proximidad.

50 La cámara externa 8 proporciona un punto de vista separado de la zona de interés (debido tanto a la ubicación física como a las diferentes propiedades de la cámara, como la resolución, los ángulos de apertura o visión, las distancias focales, etc), y no tiene ningún dato de audio directamente asociado a ella. Sin embargo, ha de tenerse en cuenta que, dado que la matriz de micrófonos 4 cuenta con una cámara integrada 6 (como se describirá con más detalle más adelante), la cámara externa 8 no es estrictamente necesaria, sino que mejora y aumenta las capacidades proporcionadas por la cámara integrada 6.

55 La matriz de micrófonos 4 está compuesta por una red bidimensional de micrófonos (aunque se apreciará que también se puede utilizar una matriz de micrófonos tridimensional). Cada micrófono de la matriz 4 proporciona un canal de audio individual, cuyo audio producido difiere ligeramente del de cualquier otro micrófono de la matriz 4. Por ejemplo, dado que están en posiciones diferentes, cada micrófono podría recibir una señal sonora de una fuente de sonido (como la de un VANT) en un momento ligeramente diferente y con diferentes fases debido a la variación de la distancia que la señal de sonido ha tenido que recorrer desde la fuente hasta el micrófono.

60 Los datos de audio procedentes de la matriz de micrófonos pueden ser analizados utilizando la generación de haces. La generación de haces se utiliza para crear una serie de canales de audio o "haces" que el procesador 10 analiza para determinar la presencia y el origen de una señal de audio de interés recibida. Si los datos de audio de un haz en

- particular son de interés, es decir, si se detecta un sonido particular, como el sonido de un dron, dentro de los datos que corresponden al haz, los ángulos que forman dicho haz proporcionan una indicación de la dirección desde la que se ha originado el sonido, porque los ángulos del haz son conocidos a priori para una dada celda espacial. El procesador es entonces capaz de determinar que el sonido se originó en un lugar a lo largo del haz en el espacio 3D, es decir, dentro de la región de la zona de interés mapeada a la celda espacial correspondiente al haz. Ha de tenerse en cuenta que la generación de haces proporciona solo la dirección desde la que se ha originado el sonido y no la distancia, aunque la distancia puede ser determinada por realizaciones de la presente invención utilizando otras técnicas, tal y como se describirá más adelante.
- La figura 3 muestra un conjunto de celdas espaciales 12 utilizadas por el procesador 10 del sistema de detección 2 de la figura 2. Como puede observarse en la figura, el procesador 10 divide la zona de interés en un conjunto de celdas espaciales 12, que en esta realización particular son celdas triangulares que encajan para formar una malla.
- Cada celda individual del conjunto 12 corresponde a un haz formado por la matriz de micrófonos 4 y, por lo tanto, el procesador puede determinar si un VANT está presente en un área determinada con una resolución tan fina como el tamaño de la malla lo permita. Si bien la malla que forma el conjunto 12 en esta realización particular está compuesta de elementos triangulares, se apreciará que la malla podría estar formada a partir de otras formas y que dichas mallas son, de por sí, conocidas en la técnica.
- Cada celda del conjunto 12 tiene asociada una puntuación de probabilidad que corresponde a la probabilidad de que un dron esté presente en dicha celda según esté determinado por el procesador 10. Dicha puntuación de probabilidad es una función de tres puntuaciones parciales, tal y como se describirá a continuación.
- La primera puntuación parcial de la que depende la puntuación de probabilidad es una puntuación de análisis de audio. La puntuación de análisis de audio se genera mediante un algoritmo de análisis de audio que compara los datos de audio correspondientes a cada celda espacial (y, por extensión, un haz de la matriz de micrófonos) con las firmas de audio de una biblioteca. Más adelante se discute en mayor detalle un posible algoritmo haciendo referencia a la figura 7, sin embargo, se apreciará que hay un número de dichos algoritmos, por ejemplo, la extracción y la selección de características, tal y como se describe en el documento FR2923043 (Orelia SAS), discutidos más adelante haciendo referencia a la figura 17, que pueden aplicarse perfectamente de acuerdo con la presente invención. Las celdas con señales de sonido que presentan una coincidencia cercana en la biblioteca de firmas de audio recibirán una puntuación de análisis de audio más alta que las celdas que no producen ninguna coincidencia cercana con respecto a ninguna firma de la biblioteca.
- Una puntuación de intensidad de audio se utiliza como segunda puntuación parcial por el procesador 10 a la hora de determinar la puntuación de probabilidad para cada celda dentro del conjunto 12. La puntuación de intensidad de audio se genera comparando la amplitud de la porción del espectro de los datos de audio correspondientes a cada celda espacial con un umbral predeterminado. Los vehículos aéreos no tripulados presentan una tendencia a producir sonidos de volumen relativamente alto, particularmente a ciertas frecuencias. Esta operación tipo umbral actúa para filtrar las fuentes de sonido de fondo que probablemente tendrán una menor amplitud en la región espectral relevante que el sonido de un VANT a detectar. Las celdas con señales de mayor amplitud espectral relevante reciben una puntuación de intensidad de audio más alta que las celdas con señales de menor amplitud espectral relevante. Las celdas con una más alta puntuación de intensidad de audio pueden recibir una alta prioridad durante el análisis de audio, lo cual implica que estas celdas de alta puntuación son analizadas en busca de firmas correspondientes a un dron antes que las celdas de menor puntuación.
- Cada celda del conjunto 12 recibe también una puntuación de análisis de vídeo que se genera mediante un algoritmo de procesamiento de imagen. Se aplica un algoritmo de procesamiento de imagen o de visión artificial a los datos de vídeo correspondientes a cada celda espacial y se analizan las propiedades características asociadas a los VANTs. Por ejemplo, el algoritmo de procesamiento de imagen puede incluir: análisis del color; Análisis de la textura; Segmentación o "clustering" de imágenes; Detección de bordes; Detección de esquinas; O cualquier otra combinación de estas y/u otras técnicas de procesamiento de imágenes que estén bien documentadas en la técnica.
- El algoritmo de procesamiento de imágenes de esta realización particular también incluye detección de movimiento. Hay una serie de algoritmos de detección de movimiento, como los que utilizan plantillas de movimiento, que están bien documentados de por sí en la técnica. Entre los algoritmos ejemplares particularmente adecuados para esta invención se incluyen OpenCV y Optical Flow.
- A continuación, se calcula una puntuación de probabilidad para cada una de las celdas partiendo de las puntuaciones obtenidas en el análisis de audio individual, de intensidad de audio y en los análisis de vídeo, y la puntuación de probabilidad se actualiza después de cada iteración de análisis de audio y clasificación. Existen diversas formas diferentes de calcular dicha puntuación de probabilidad. Por ejemplo, la puntuación de probabilidad puede ser un total de las múltiples puntuaciones parciales o puede ser un cálculo medio de las mismas. De forma alternativa, la puntuación de probabilidad podría ser una media ponderada en la que las diferentes puntuaciones parciales reciben diferentes ponderaciones que pueden ser configuradas por el diseñador o modificadas de forma dinámica por el procesador 10.

El conjunto de celdas 12 forma un "mapa de calor" de probabilidad, en el que la probabilidad de que un VANT esté presente en cualquier punto dado dentro de la proyección en 2D de la zona de interés en 3D se representa en forma de mapa.

5 La figura 4 muestra el vehículo aéreo no tripulado 50 de la figura 1 habiendo entrado en la zona de interés del sistema de detección 2 de la figura 2. El VANT 50 pasa a ser visible para la matriz de micrófonos 4, su cámara integrada 6 y la cámara externa 8. Tal y como se puede observar en la figura 5, el VANT 50 ocupa varias celdas 12.

10 La figura 6 muestra el mapa de probabilidad de detección espacial después del análisis realizado por el procesador 10. Un subconjunto de celdas 14 ocupadas por el VANT 50 aparece sombreado para indicar que sus puntuaciones de probabilidad correspondientes son altas en comparación con el resto de las celdas 12. Dicho sombreado indica que el procesador 10, tras haber llevado a cabo los análisis de audio y vídeo descritos anteriormente, ha calculado que las puntuaciones de probabilidad en este subconjunto 14 son mayores que las de las celdas circundantes 12.

15 En este ejemplo particular, las puntuaciones de probabilidad en cada celda del subconjunto 14 son superiores al umbral de detección aplicado por el procesador 10. Por consiguiente, el sistema de detección 2 determina que el VANT 50 está ubicado en el espacio aéreo que corresponde a las ubicaciones reales a las que está mapeado el subconjunto de celdas 14. El sistema de detección 2 puede, entonces, activar una alarma para alertar al usuario de que el VANT 20 50 ha sido detectado. El sistema de detección 2 también podría comenzar a rastrear los movimientos del VANT 50.

La figura 7 muestra un ejemplo de un proceso de análisis de audio en el que se utiliza una biblioteca de firmas de audio 80. El procesador 10 analiza los datos procedentes de la matriz de micrófonos 4 para determinar si los sonidos que se están recibiendo corresponden a un VANT y, de ser así, qué modelo de VANT es más probable que sea.

25 Los datos de audio procedentes de la matriz de micrófonos 4 son expuestos a transformada de Fourier para producir un espectro de frecuencia 70 correspondiente a los datos de audio recibidos para una celda determinada del conjunto de celdas 12 (es decir, el audio correspondiente a un haz particular). Dicho espectro de frecuencia 70 muestra la magnitud  $|A|$  para cada frecuencia  $f$  dentro de un rango específico. En este ejemplo particular, el rango es de 100 Hz a 10 kHz. Si bien el espectro de frecuencia 70 mostrado aquí parece ser continuo, los espectros serán, por lo general, discretos en aplicaciones reales, debido a los niveles de cuantificación finita utilizados por el procesador 10. Se entenderá que también sería posible aplicar otras transformadas de dominio relacionadas con la transformada de Fourier conocidas en la técnica por sí mismas, como la transformada de coseno discreta (DCT) o la transformada del coseno discreta modificada (MDCT), para producir un espectro de frecuencia adecuado.

35 Este espectro de frecuencia 70 se compara, a continuación, con las firmas de audio de una biblioteca 80 para buscar una coincidencia. En aras de la claridad, tan solo se muestran tres firmas de audio 72, 74, 76 en la figura; Sin embargo, un sistema práctico tendrá, por supuesto, una biblioteca mucho más amplia. El procesador 10 determina que el espectro 70 no es una coincidencia cercana para los espectros asociados con dos de las firmas de audio 72, 76, pero 40 sí coincide con los espectros de la firma de audio 74 central señalada en la figura con una marca de verificación. De esta forma, el procesador determina a través del análisis de audio que el espectro 70 de la celda asociada corresponde no solo a la presencia del VANT 50, sino que también indica qué tipo de VANT es.

45 La figura 8 muestra el conjunto de celdas espaciales 12' de la figura 5 habiendo sido refinadas tras la detección del vehículo aéreo no tripulado 50. Si bien, haciendo referencia a la figura 5, se describió anteriormente que las celdas 14 contaban con una puntuación de probabilidad asociada que excedía el umbral de detección, podría darse el caso de que, a pesar de ser más alto de lo normal, la puntuación no fuera suficiente como para afirmar con una certeza razonable que el VANT 50 estaba presente en la zona de interés.

50 De forma alternativa, el procesador 10 puede estar razonablemente seguro de que el VANT 50 está en la zona de interés y pasar a desear obtener una mejor estimación de su posición y de sus dimensiones.

55 En ambos casos, podría darse la situación de que la puntuación de probabilidad asociada a dichas celdas 14 exceda el umbral de cambio de resolución. Una vez que esto ocurre, el procesador puede tomar la decisión de aumentar la resolución de la malla, produciendo así un conjunto refinado de celdas 12'. Como es posible observar al comparar el conjunto de celdas 12' de la figura 8 con el conjunto de celdas 12 de la figura 5, las celdas triangulares se han hecho más pequeñas y numerosas, es decir, la densidad de las celdas ha aumentado. Por ejemplo, los haces formados utilizando la matriz de micrófonos 4 podrían haber estado separados por espacios angulares de  $10^\circ$  y ahora, estar separados por  $1^\circ$ . Esto es algo que puede efectuarse en toda la zona de interés o, preferiblemente, solo en las proximidades de la puntuación de probabilidad aumentada.

60 Ahora que las celdas individuales son más pequeñas, lo cual, por supuesto, aumenta los requisitos de potencia de procesamiento, el subconjunto de celdas 14' que corresponde a la posición del VANT 50 proporciona un "mayor ajuste" a la forma del VANT 50. El aumento de la densidad sombreada también indica que la puntuación de probabilidad asociada con cada una de las celdas del subconjunto 14' es más alta de lo que era anteriormente en la figura 5, es decir, el procesador 10 ahora está más seguro de que el VANT 50 está verdaderamente presente en esa área.

65

La figura 9 muestra un sistema de detección de vehículos aéreos no tripulados 2 de acuerdo con una realización adicional de la presente invención, que utiliza múltiples matrices de micrófonos 4, 16. En esta realización, el sistema 2, tal y como se ha descrito previamente, es provisto de una matriz de micrófonos adicional 16. Esta matriz de micrófonos 16 particular no posee una cámara integrada como la matriz original 4, pero aquellos expertos en la técnica apreciarán que cualquier combinación de matrices, con o sin cámaras integradas, puede ser añadida al sistema 2 según lo requiera una aplicación en cuestión.

En dicho caso, las dos matrices de micrófonos 4, 16, pueden utilizarse cada una de ellas en un proceso de generación de haces y cada una de ellas proporciona datos de audio al procesador 10. Las matrices de micrófonos 4, 16 pueden proporcionar diferentes "puntos de vista" de la zona de interés. Ello permite la monitorización de diferentes "subzonas" de la zona de interés por parte de cada matriz 4, 16, ya que cada matriz tan solo puede proporcionar una vista de un área finita.

De forma alternativa, si las dos matrices 4, 16 son posicionadas lo suficientemente juntas, pueden combinarse para proporcionar la funcionalidad de una sola y más grande "supermatriz". Dicha supermatriz tiene, entonces, una mayor resolución que una matriz individual.

La figura 10 muestra un sistema de detección de vehículos aéreos no tripulados 2 de acuerdo con una realización adicional de la presente invención, que utiliza múltiples cámaras externas 8, 18. De forma similar a la realización descrita con referencia a la figura 9, los expertos en la técnica apreciarán que podrá contemplarse cualquier combinación de cámaras externas, matrices de micrófonos con cámaras integradas y matrices de micrófonos sin cámaras integradas.

Las dos cámaras externas 8, 18 son posicionadas en diferentes ubicaciones y cada una proporciona una vista diferente de la zona de interés, tal y como se describirá a continuación haciendo referencia a la figura 11. Las dos cámaras 8, 18 pueden contar con propiedades diferentes, como diferentes distancias focales, capacidades de zoom, habilidad para efectuar paneos (pan) y/o inclinarse (tilt), etc. o pueden ser idénticas, dependiendo de los requisitos de la aplicación.

Cada cámara puede estar representada por sus parámetros intrínsecos, como se muestra a continuación haciendo referencia a la ecuación 1:

$$A_n = \begin{bmatrix} \alpha_{x,n} & \gamma_n & u_{0,n} \\ 0 & \alpha_{y,n} & v_{0,n} \\ 0 & 0 & 1 \end{bmatrix}$$

Ecuación 1: Parámetros intrínsecos de la cámara

en donde:  $A_n$  es la matriz de parámetros intrínsecos de la cámara  $n^{th}$ ;  $\alpha_{x,n}$  es la distancia focal multiplicada por un factor de escala en la dirección x- para la cámara  $n^{th}$ ;  $\alpha_{y,n}$  es la distancia focal multiplicada por un factor de escala en la dirección y- para la cámara  $n^{th}$ ;  $\gamma_n$  es un parámetro sesgado de la cámara  $n^{th}$ ; y  $u_{0,n}, v_{0,n}$  es el "punto fundamental" de la imagen producida por la cámara  $n^{th}$ , que es, generalmente pero no siempre, el centro de la imagen en coordenadas de píxeles. Se apreciará que este es un modelo de los parámetros intrínsecos de la cámara y que se pueden incluir otros parámetros en la matriz de parámetros intrínsecos, como la distorsión óptica, proporcionando, por ejemplo, la distorsión de barril, la distorsión de cojín, la distorsión de bigote, etc.

La figura 11 muestra los puntos de vista 20, 22 de las cámaras 8, 18 respectivamente, tal y como se ha descrito anteriormente haciendo referencia a la figura 10. La primera cámara 8 proporciona un primer punto de vista 20 de la zona de interés, que tiene una cierta "rotación" y un cierto "sesgo" asociados con sí mismo, debido a la posición y al ángulo en los que está instalada la cámara 8. De forma similar, la segunda cámara 18 proporciona un segundo punto de vista 22 de la zona de interés que presenta una rotación y un sesgo diferentes a los del primer punto de vista 20. Cada cámara 8, 18, por lo tanto, tiene vistas de la zona de interés ligeramente diferentes (por ejemplo, la segunda cámara 18 no puede "ver" la nube situada más a la izquierda, pero la primera cámara 8 sí).

La figura 12 muestra el corregistro de los puntos de vista 20, 22, tal y como se ha descrito anteriormente haciendo referencia a la figura 11. Como se puede observar en la figura 12, existe un área 21 dentro del primer punto de vista 20 que posee una fuerte correspondencia con un área 23 dentro del segundo punto de vista 22. Dado que las posiciones y las propiedades de las cámaras 8, 18 son conocidas, estos puntos de vista 20, 22 pueden ser comparados directamente mapeándolos entre sí. De hecho, incluso aunque las posiciones relativas de la cámara no se conocieran

a priori, existen numerosas técnicas de procesamiento de imagen conocidas en la técnica por sí mismas que podrían determinar el mapeado de cámara a cámara.

5 Con dicho conocimiento, los dos puntos de vista 20, 22 pueden ser corregistrados y también traducidos a una imagen del "mundo real" con profundidad. Las dos áreas 21, 23, por ejemplo, pueden ser mapeadas de nuevo a un área del mundo real 24 que "mira" de frente hacia la zona de interés.

10 Ello se consigue al tener una matriz C que representa la posición o la "pose" de la cámara tal y como se indica en la ecuación 2 mostrada a continuación:

$$C_n = [-R_n^T T_n]$$

*Ecuación 2: Parámetros extrínsecos de la cámara*

15 en donde  $C_n$  es la matriz de pose de la cámara  $n^{\text{th}}$ ;  $R_n$  es una matriz de rotación para la cámara  $n^{\text{th}}$  que traduce la rotación de la cámara a las coordenadas comunes; Y  $T_n$  es una matriz de traducción para la cámara  $n^{\text{th}}$  que traduce la posición de la cámara a las coordenadas comunes, donde la forma general de la matriz de rotación  $R_n$  y la matriz de traducción  $T_n$  son conocidas en el arte por sí mismas.

20 El mapeo de las coordenadas locales de la cámara al sistema de coordenadas común se puede conseguir utilizando los ángulos de Euler o los ángulos de Tait-Bryan para rotar las coordenadas locales al sistema de coordenadas común, en donde las rotaciones se producen alrededor de los ejes x-, y- y z-. En un ejemplo, se utiliza un sistema de coordenadas diestro, por ejemplo, el eje x- es positivo hacia el lado derecho, el eje y- es positivo hacia abajo y el eje z- es positivo a lo largo de la línea de visión. Ello implica llevar a cabo cuatro rotaciones diferentes, cada una de las cuales se puede representar como una matriz de rotación separada y dichas cuatro matrices de rotación pueden ser combinadas en una sola matriz de rotación individual que proporciona:

1. Una rotación fija de 270° alrededor del eje x- de la cámara;
- 30 2. Pan: rotación alrededor del eje y- de la cámara;
3. Tilt: rotación alrededor del eje x- de la cámara; Y
- 35 4. Roll: rotación alrededor del eje z- de la cámara.

Por consiguiente, el sistema de coordenadas de la cámara puede alinearse con el sistema de coordenadas del mundo real. En el caso del sistema de coordenadas universal transversal de Mercator (UTM), ello implica que el eje x- de la cámara esté alineado con el este, el eje -y de la cámara esté alineado con el norte y el eje z- de la cámara esté alineado con la altura.

40 Las posiciones y los ángulos correspondientes a la/las matriz/matrices de micrófonos pueden ser mapeados a las coordenadas comunes de una forma similar y, por lo tanto, todas las fuentes de datos de audio y vídeo pueden utilizar un sistema de coordenadas común, que también es utilizado por el procesador 10 como base para el mapa de probabilidad que comprende el conjunto de celdas 12, 12'.

45 Dado que hay múltiples cámaras 8, 18 con un área superpuesta 24 y que se conoce la relación existente entre dichas cámaras 8, 18, es posible determinar la profundidad de un objeto como el VANT 50 dentro de dicha área 24 al comparar los píxeles de cada imagen correspondiente al VANT 50 en los dos puntos de vista 20, 22, para lo cual se utilizan técnicas estereoscópicas que son conocidas en la técnica por sí mismas. Entre la cámara integrada 6 y una o ambas cámaras externas 8, 18, es posible llevar a cabo un emparejamiento similar para proporcionar más información sobre la profundidad. Dicha información sobre la profundidad puede ser ampliada por los datos de audio procedentes de la matriz de micrófonos 4.

50 La figura 13 muestra la operación de un dron de calibración 90 utilizado para mapear las celdas espaciales a coordenadas GPS del mundo real. Se hace volar el dron de calibración 90 a través de la zona de interés objeto de monitorización por parte del sistema de detección 2. Un usuario hace volar al dron de calibración (ya sea de forma manual o utilizando una ruta de vuelo automática predeterminada) a lo largo de la ruta 94.

55 El dron de calibración también está equipado con un sensor GPS (sistema de posicionamiento global) 92. El sensor GPS 92 se utiliza para registrar las coordenadas del mundo real del dron de calibración conforme este se desplaza a lo largo de la ruta 94. El procesador 10 cuenta con una marca de tiempo común compartida con el sensor GPS 92 y,

por consiguiente, los datos GPS registrados por el dron de calibración 90 pueden compararse directamente con los datos de audio y vídeo proporcionados por la matriz de micrófonos 4, la cámara integrada 6 y la cámara externa 8. Esto permite establecer una correspondencia entre las celdas espaciales y las coordenadas GPS, tal y como se describirá a continuación.

5 La figura 14 muestra cómo se calibran las celdas espaciales 12 utilizadas por el procesador de la figura 13 utilizando el dron de calibración 90. Dado que el sensor GPS 92 y el procesador 10 están sincronizados en tiempo, el procesador puede comparar los tiempos en los que el dron de calibración 90 ha atravesado cada celda con los datos GPS procedentes del sensor GPS 92 y obtener así un mapeo de calibración de relación uno a uno entre las celdas espaciales 12 y las coordenadas GPS del mundo real. Entonces, durante el funcionamiento habitual, un VANT detectado, como el VANT 50, puede ser identificado en un mapa del mundo real, puesto que las celdas en las que se detecta poseen una posición conocida. Ello puede lograrse traduciendo las coordenadas a coordenadas del sistema de coordenadas universal transversal de Mercator (UTM) correctas. Las coordenadas podrían, por supuesto, traducirse a otros sistemas de coordenadas según lo requiera el usuario final.

15 La figura 15 muestra una fuente de ruido constante que puede ser compensada por el sistema de detección 2. En esta figura, el sistema de detección 2 ha sido instalado en una ubicación próxima a un aerogenerador 100. El aerogenerador 100, cuando está en funcionamiento, produce un ruido relativamente constante, que podría causar dificultades en la detección del vehículo aéreo no tripulado por sonido. No obstante, el procesador 10 está dispuesto de manera que puede calibrarse para ignorar tales fuentes de ruido constante. Ello puede lograrse calibrando el sistema cuando no hay drones en el área, de manera que cualquier sonido que se escuche durante la calibración que luego se escuche durante el funcionamiento del sistema podrá ser sustraído del sonido percibido durante el funcionamiento. Dicho procedimiento de filtrado podría implicar la cancelación espacial utilizando algoritmos de generación de haces, procedimientos de filtrado en el dominio de la frecuencia o una combinación de ambos. De forma adicional o alternativa, el procesador 10 puede ser calibrado para ignorar ciertas frecuencias de sonido de las que se sabe que son fuentes de ruido, por ejemplo, el ruido constante de 50 Hz que produce el aerogenerador 100, o para eliminar espacialmente la banda de la posición conocida y estacionaria del ruido no deseado.

30 La figura 16 muestra un subconjunto de celdas espaciales 14" de la figura 8 que han sido refinadas adicionalmente en las proximidades del vehículo aéreo no tripulado detectado. En este ejemplo particular, el procesador ha decidido aumentar de forma adicional la resolución de la malla solo en las proximidades del VANT 50, produciendo por tanto un conjunto de celdas refinadas 14". Tal y como se puede observar al comparar el conjunto de celdas 14' de la figura 8 con el conjunto de celdas 14" de la figura 16, las celdas triangulares se han hecho aún más pequeñas, es decir, la densidad de las celdas ha aumentado aún más. Este nuevo subconjunto de celdas 14" proporciona un ajuste aún mayor a la forma del VANT 50. Se apreciará que podría no haber un paso intermedio consistente en aumentar la resolución de forma global antes de aumentarla solo en las proximidades del dron, y que la resolución podría ser solo aumentada de forma local en función de los requisitos de procesamiento.

40 La figura 17 muestra un diagrama de bloques de un ejemplo adicional de un proceso de análisis de audio que utiliza un algoritmo de detección y clasificación de características 200. En este algoritmo 200, los datos de audio 202 correspondientes a un haz particular se pasan a través de un bloque de extracción de características 204, un bloque de selección de características 206 y un bloque de clasificación 208 para determinar la clasificación 212 de los datos de audio 202.

45 El bloque de extracción de características 204 implementa un análisis temporal utilizando la forma de onda de la señal de audio 202 y/o un análisis espectral utilizando la representación espectral de la señal de audio 202 para llevar a cabo el análisis. El bloque de extracción de características 204 analiza pequeños segmentos de la señal de audio 202 a la vez y busca características concretas, como puedan ser el tono, el timbre, el roll-off, la tasa de cruces por cero, el centroide, el flujo, la intensidad del ritmo, la regularidad rítmica, la tasa de distorsión armónica, etc.

50 El conjunto de características 205 extraído por el bloque de extracción de características 204 es entonces introducido en el bloque de selección de características 206. A continuación, el bloque de selección de características 206 selecciona un subconjunto específico de características 207 que son elegidas para ser las más indicativas de la fuente de ruido que se ha de buscar (por ejemplo, un dron). El subconjunto de características 207 es elegido para proporcionar un nivel aceptable de rendimiento y un alto grado de precisión al proceso de clasificación (por ejemplo, no proporciona demasiados falsos positivos y falsos negativos) y reduce la complejidad computacional al garantizar que las características elegidas no sean redundantes, es decir, cada una de las características elegidas en el subconjunto 207 proporciona información adicional de utilidad para la clasificación que no ha sido proporcionada ya por otra característica del subconjunto 207.

60 El subconjunto de características elegido 207 se pasa luego al bloque de clasificación 208. El bloque de clasificación 208 utiliza, entonces, un algoritmo de clasificación, como un clasificador de los  $k$  vecinos más cercanos o un clasificador de mezcla de gaussianas. El bloque clasificador 208 puede tomar también modelos estadísticos 210 como input. Dichos modelos estadísticos 210 pueden haberse construido en función de unos datos de entrenamiento en los que las etiquetas de clasificación (por ejemplo, un modelo específico de dron) son asignadas manualmente a los datos de audio correspondientes y pueden ayudar al bloque de clasificación 208 a la hora de determinar qué está presente

dentro de la señal de audio 202. El bloque de clasificación 208, entonces, emite una etiqueta de clasificación 212 como, por ejemplo, "dron presente", "dron no presente", o puede nombrar un modelo específico de dron.

5 Por lo tanto, se observará que en el presente documento se describe un sistema distribuido y colaborativo de matrices de micrófonos y cámaras que utiliza varios análisis estadísticos, filtros espaciales y algoritmos de filtrado de frecuencia de tiempo para detectar, clasificar y rastrear vehículos aéreos no tripulados en un área potencialmente grande, en varios entornos diferentes. Si bien las realizaciones particulares se han descrito en detalle, los expertos en la técnica apreciarán que es posible implementar muchas variaciones y modificaciones utilizando los principios de la invención presentada en este documento.

10

**REIVINDICACIONES**

1. Un sistema (2) para la detección, la clasificación y el rastreo de vehículos aéreos no tripulados (50) en una zona de interés, comprendiendo el sistema:

al menos una matriz de micrófonos (4, 16) que incluye varios micrófonos, estando la, al menos, una matriz de micrófonos dispuesta para proporcionar datos de audio;

al menos una cámara (8, 18) dispuesta para proporcionar datos de vídeo; Y

al menos un procesador (10) dispuesto para procesar los datos de audio y los datos de vídeo para generar un mapa de probabilidad de detección espacial que comprende un conjunto de celdas espaciales (12), en donde el procesador (10) asigna una puntuación de probabilidad a cada celda existente dentro del conjunto de celdas espaciales (12), siendo dicha puntuación de probabilidad una función de:

una puntuación de análisis de audios generada por un algoritmo de análisis de audios, comprendiendo dicho algoritmo de análisis de audios la comparación de los datos de audio correspondientes a la celda espacial con una biblioteca de firmas de audio (80);

una puntuación de intensidad de audio generada mediante la evaluación de una amplitud de, al menos, una porción de un espectro (70) de los datos de audio correspondientes a la celda espacial;

Y una puntuación de análisis de vídeo generada utilizando un algoritmo de procesamiento de imágenes para analizar los datos de vídeo correspondientes a la celda espacial,

En donde el sistema está dispuesto para indicar que un vehículo aéreo no tripulado (50) ha sido detectado en una o más celdas espaciales (14) dentro de la zona de interés si la puntuación de probabilidad asignada a dicha o dichas celdas espaciales (14) excede un umbral de detección predeterminado.

2. El sistema (2) según se reivindica en cualquiera de las reivindicaciones anteriores, comprendiendo varias cámaras (8, 18) y en donde los datos de audio procedentes de la, al menos, una matriz de micrófonos (4) se utilizan para mejorar la detección de profundidad llevada a cabo utilizando las varias cámaras (8, 18).

3. El sistema (2) según cualquiera de las reivindicaciones anteriores, comprendiendo varias matrices de micrófonos (4, 16) en donde cada matriz de micrófonos incluye una cámara (6).

4. El sistema (2) según se reivindica en cualquiera de las reivindicaciones anteriores, en donde al menos dos matrices de micrófonos (4, 16) y/o cámaras (8, 18) están asociadas entre sí utilizando una relación espacial conocida entre las ubicaciones físicas de la/las matriz/matrices de micrófonos (4, 16) y/o la(s) cámara(s) (8, 18), de manera que la/las matriz/matrices de micrófonos (4, 16) y/o la(s) cámara(s) (8, 18) comparten un sistema de coordenadas común.

5. El sistema (2) según se reivindica en cualquiera de las reivindicaciones anteriores, en donde el conjunto de celdas (12) se genera de forma automática.

6. El sistema (2) según se reivindica en cualquiera de las reivindicaciones anteriores, en donde el procesador (10) está dispuesto de forma selectiva para aumentar un número de celdas espaciales (12') en, al menos, un subconjunto (14) de dicha zona de interés si la puntuación de probabilidad asignada a una o más de una celda espacial en dicho subconjunto (14) llegara a exceder un umbral de cambio de densidad de celda predeterminado, en donde, preferiblemente, el umbral de cambio de densidad de celda es inferior al umbral de detección.

7. El sistema (2) según se reivindica en cualquiera de las reivindicaciones anteriores, en donde el procesador (10) está dispuesto de forma selectiva para refinar la resolución de, al menos, una matriz de micrófonos (4, 16) y/o cámara (8, 18) si la puntuación de probabilidad asignada a dicha una o más celdas espaciales (12) excede un umbral de cambio de resolución predeterminado, en donde, preferiblemente, el umbral de cambio de resolución es inferior al umbral de detección.

8. El sistema (2) según se reivindica en cualquiera de las reivindicaciones anteriores, en donde al menos una cámara (8, 18) está dispuesta para hacer zoom en un área dentro de la zona de interés si la puntuación de probabilidad asignada a dicha o a dichas celdas espaciales (12) excede un umbral de zoom predeterminado, en donde, preferiblemente, el umbral de cambio de zoom es inferior al umbral de detección.

9. El sistema (2) según se reivindica en cualquiera de las reivindicaciones anteriores, en donde el conjunto de celdas espaciales (12) está mapeado adicionalmente a datos de calibración que comprenden varias coordenadas del sistema de posicionamiento global.

5 10. El sistema (2) según se reivindica en cualquiera de las reivindicaciones anteriores, en donde los datos de audio procedentes de al menos una matriz de micrófonos (4, 16) se utilizan para guiar el análisis de los datos de vídeo de al menos una cámara (8, 18) y/o en donde los datos de vídeo procedentes de al menos una cámara (8, 18) se utilizan para guiar el análisis de los datos de audio de al menos una matriz de micrófonos (4, 16).

11. El sistema (2) según se reivindica en cualquiera de las reivindicaciones anteriores, en donde el algoritmo de procesamiento de imágenes comprende:

10 el cálculo de un marco promedio a partir de un subconjunto de cuadros de datos de vídeo recibidos previamente;

La resta de dicho marco promedio de los cuadros de los datos de vídeo recibidos posteriormente para generar una imagen de diferencia; Y

15 la comparación de dicha imagen de diferencia con un umbral dentro de cada celda espacial visual para generar una puntuación para el análisis de vídeo.

20 12. El sistema (2) según se reivindica en cualquiera de las reivindicaciones anteriores, en donde el algoritmo de análisis de audio comprende además compensar una fuente de ruido predeterminada próxima a la zona de interés, en donde, preferiblemente, el algoritmo de análisis de audio comprende compensar la fuente de ruido predeterminada de forma automática.

25 13. El sistema (2) según se reivindica en cualquiera de las reivindicaciones anteriores, en donde el algoritmo de análisis de audio comprende un algoritmo de gradiente, en donde el algoritmo de gradiente está dispuesto para medir un cambio relativo en una distribución audio-espacial a través de una o más de una celda espacial (12).

30 14. El sistema (2) según se reivindica en cualquiera de las reivindicaciones anteriores, en donde el procesador (10) está dispuesto para procesar dichos datos de audio y visuales en una serie de períodos de tiempo repetitivos, de manera que procesa datos para cada celda espacial dentro de cada período de tiempo.

35 15. El sistema (2) según se reivindica en cualquiera de las reivindicaciones anteriores, en donde la función de la puntuación de probabilidad se modifica de forma dinámica durante el funcionamiento regular del sistema.

35

40

45

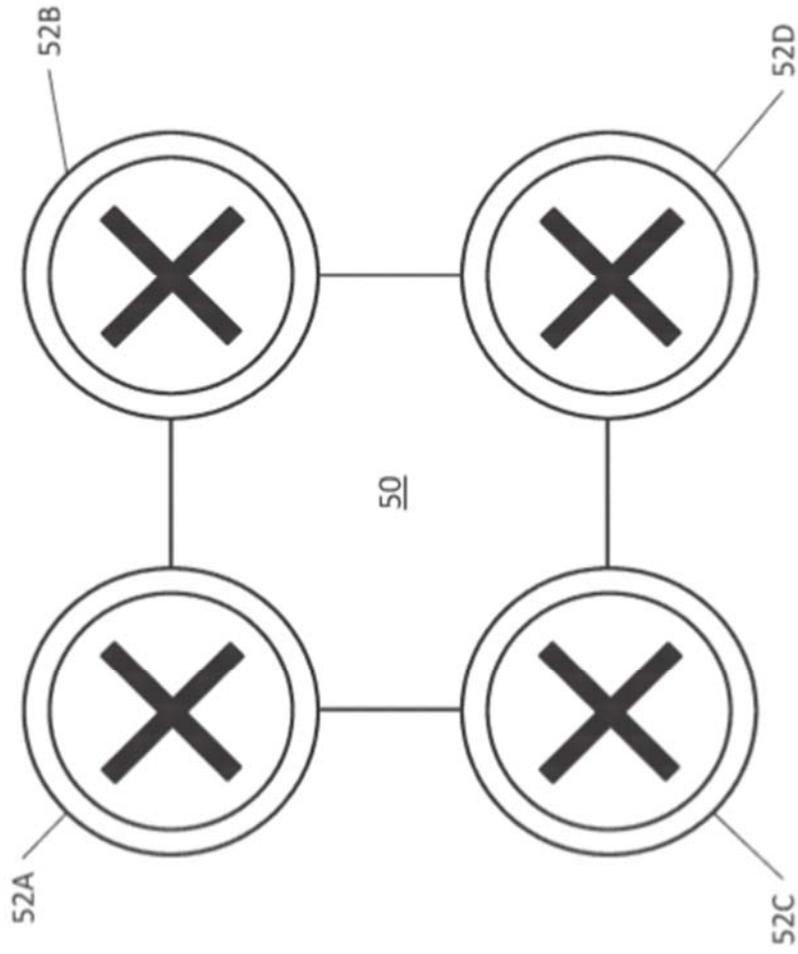


Fig. 1

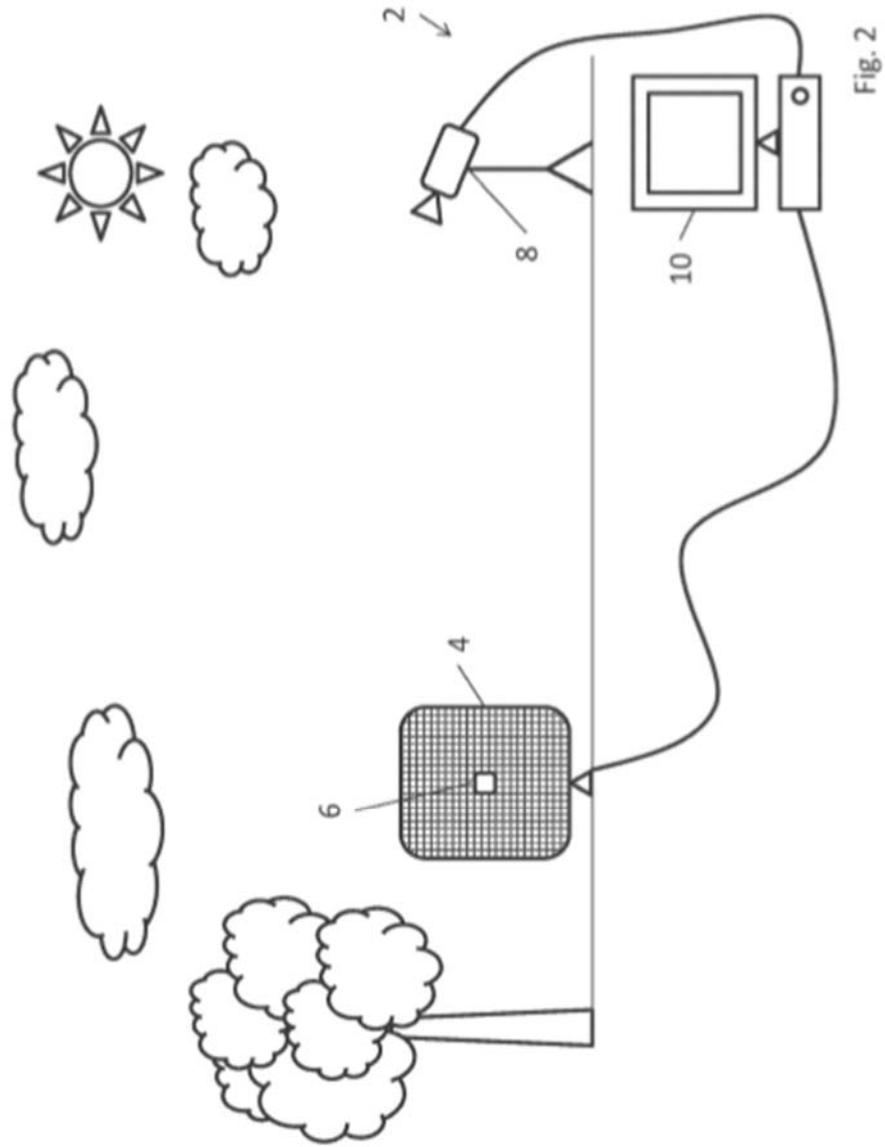


Fig. 2

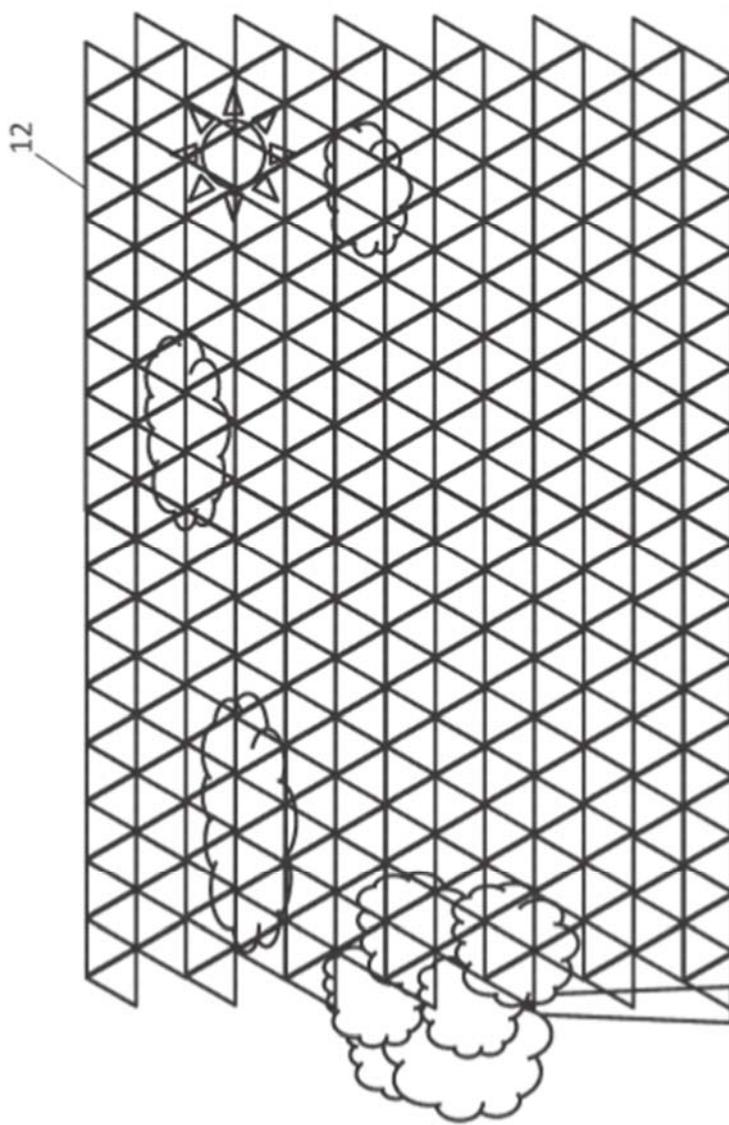


Fig. 3

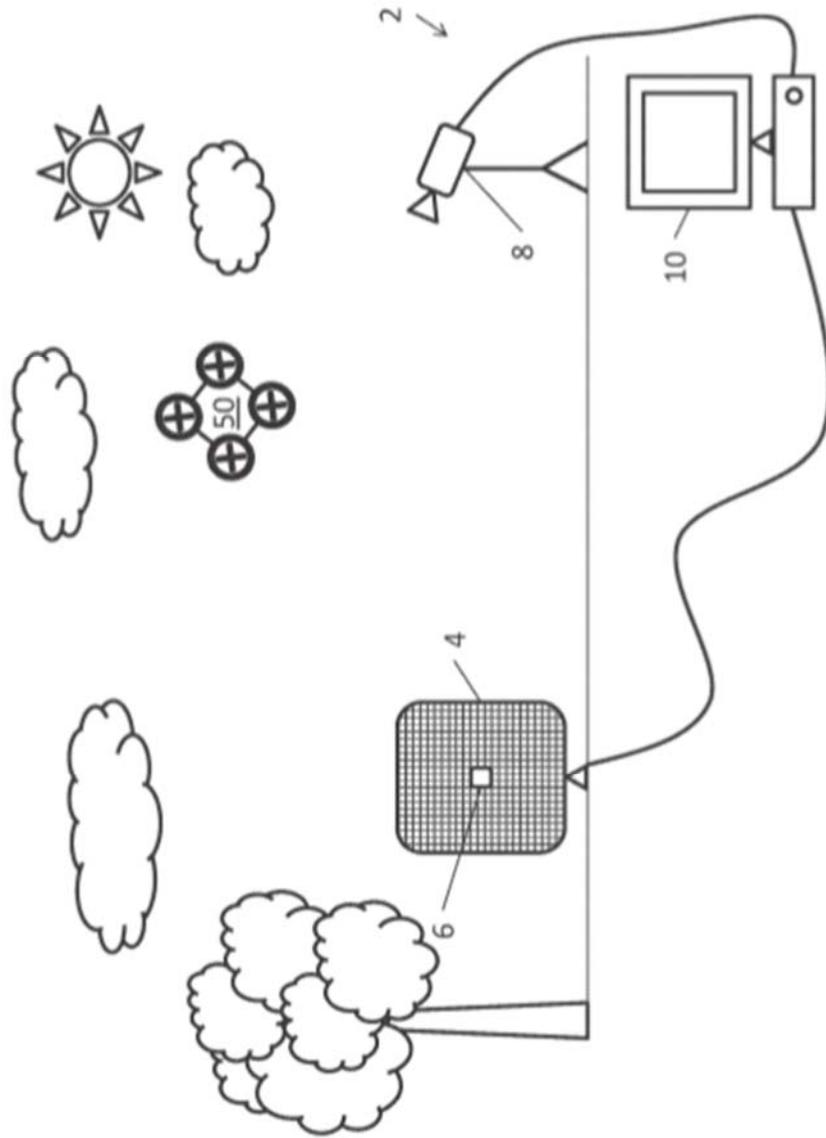


Fig. 4

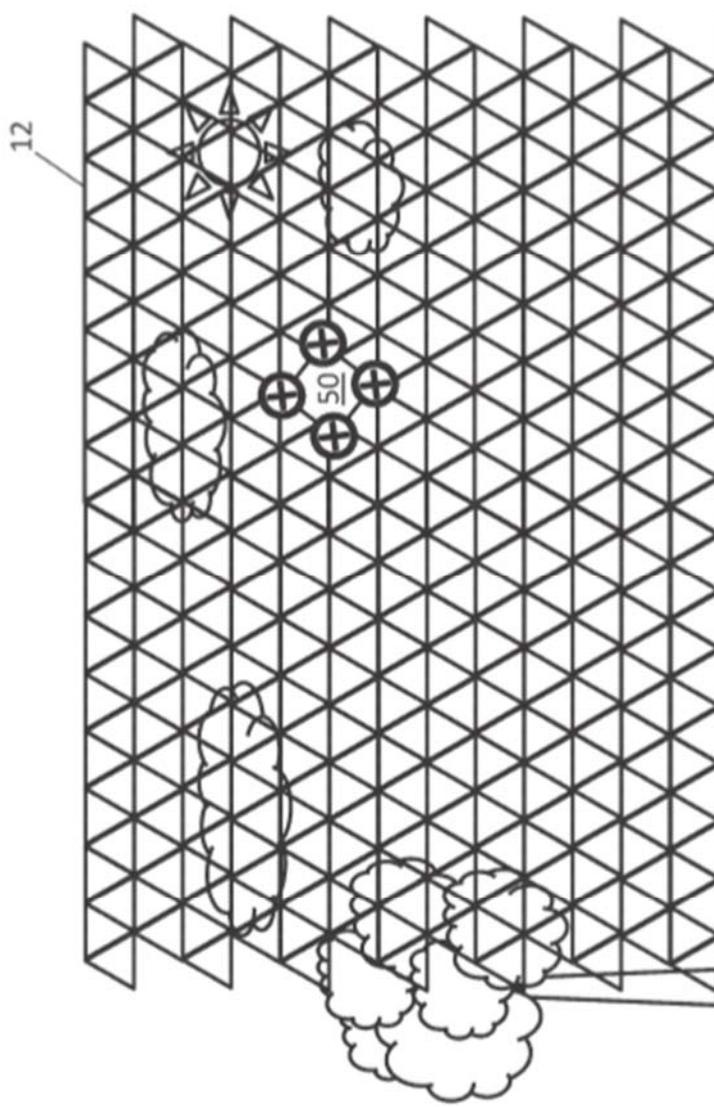


Fig. 5

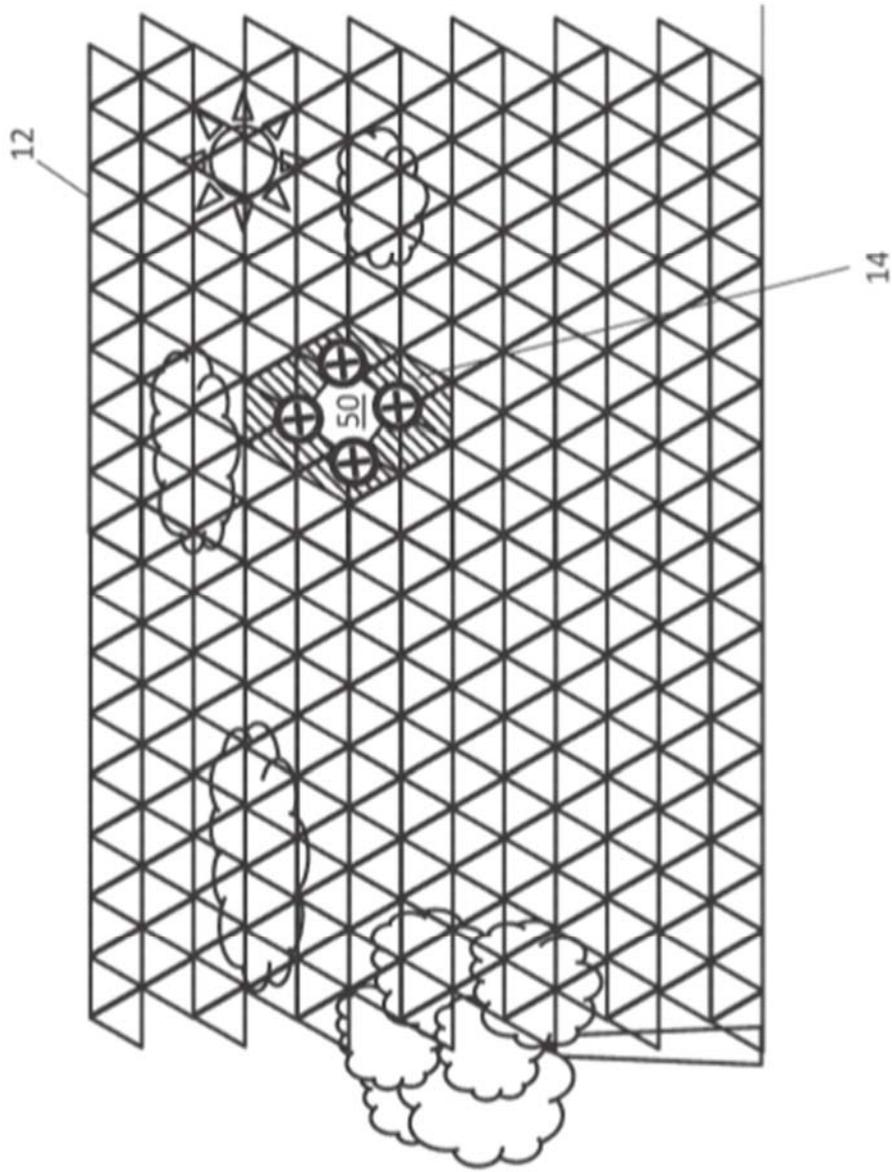


Fig. 6

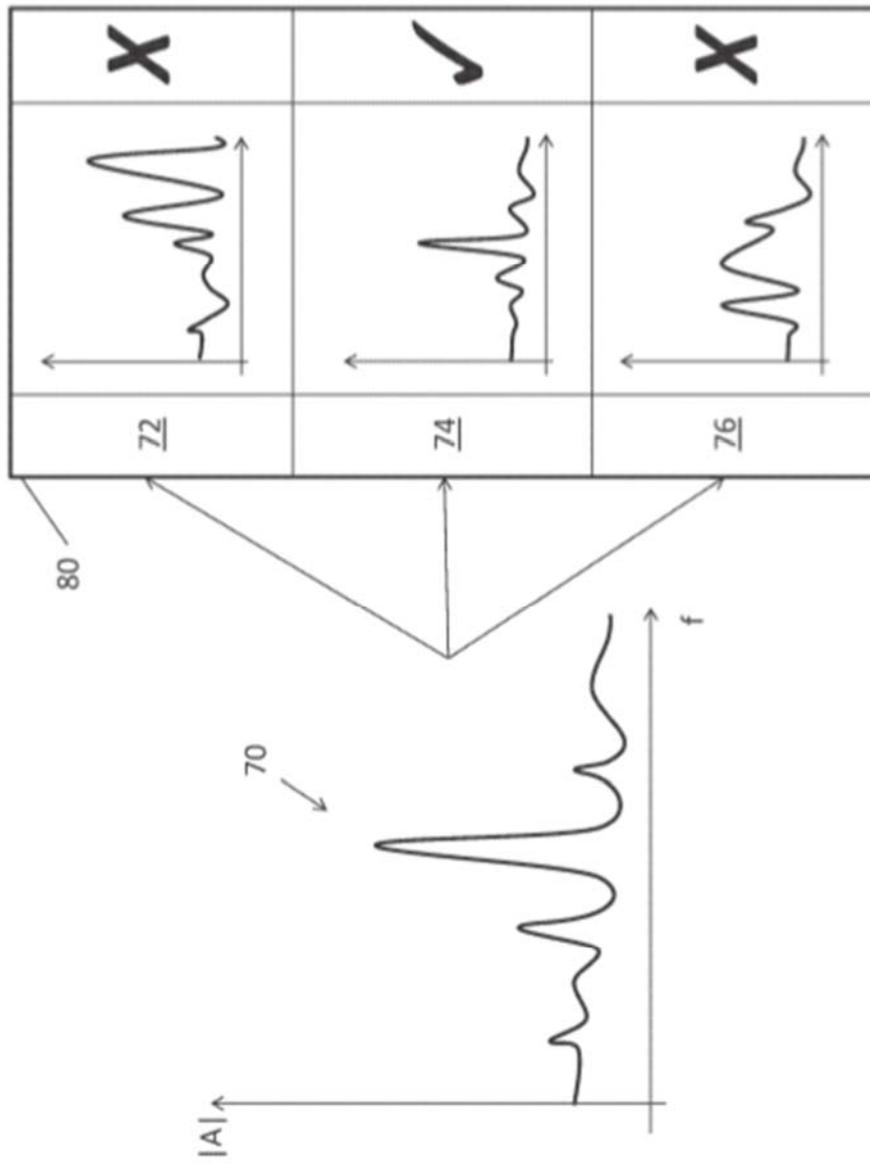


Fig. 7

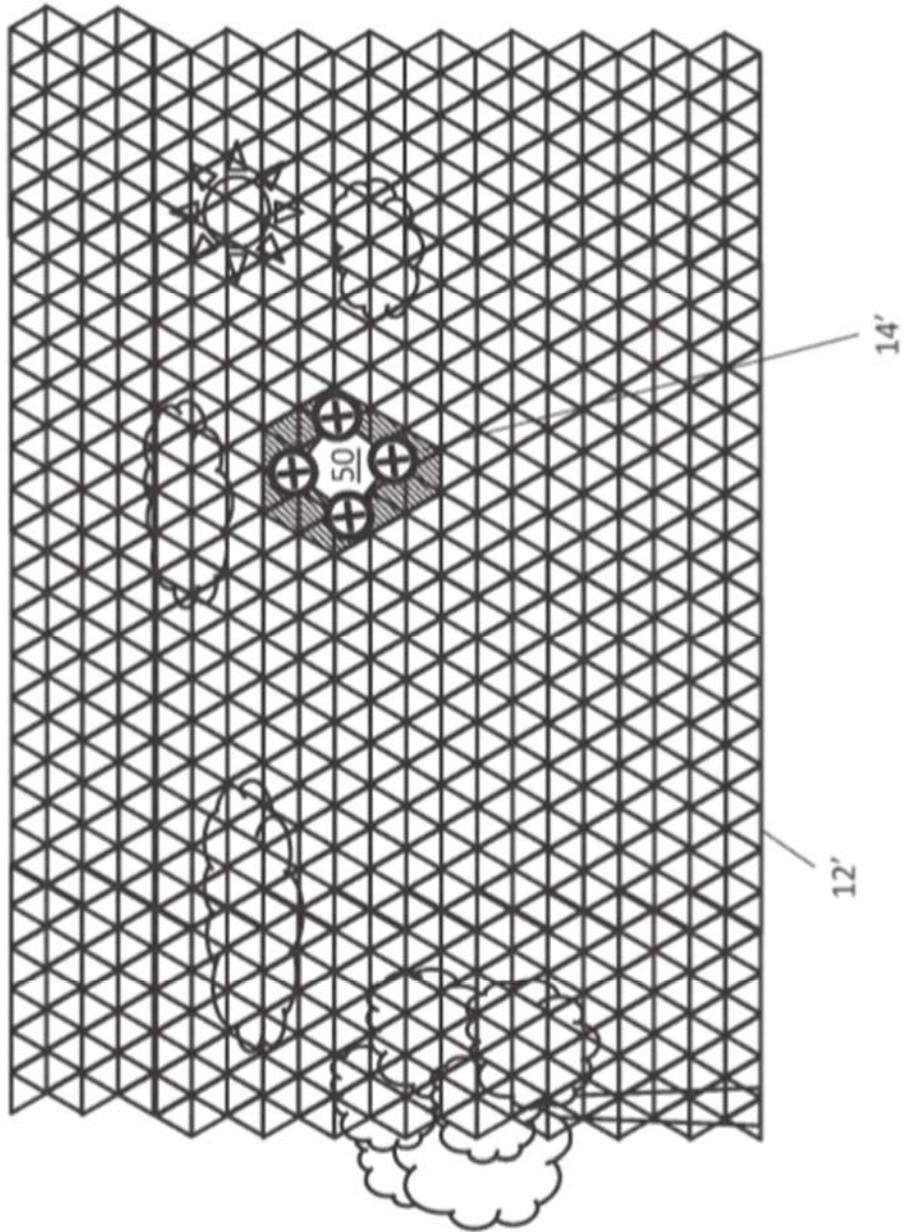


Fig. 8

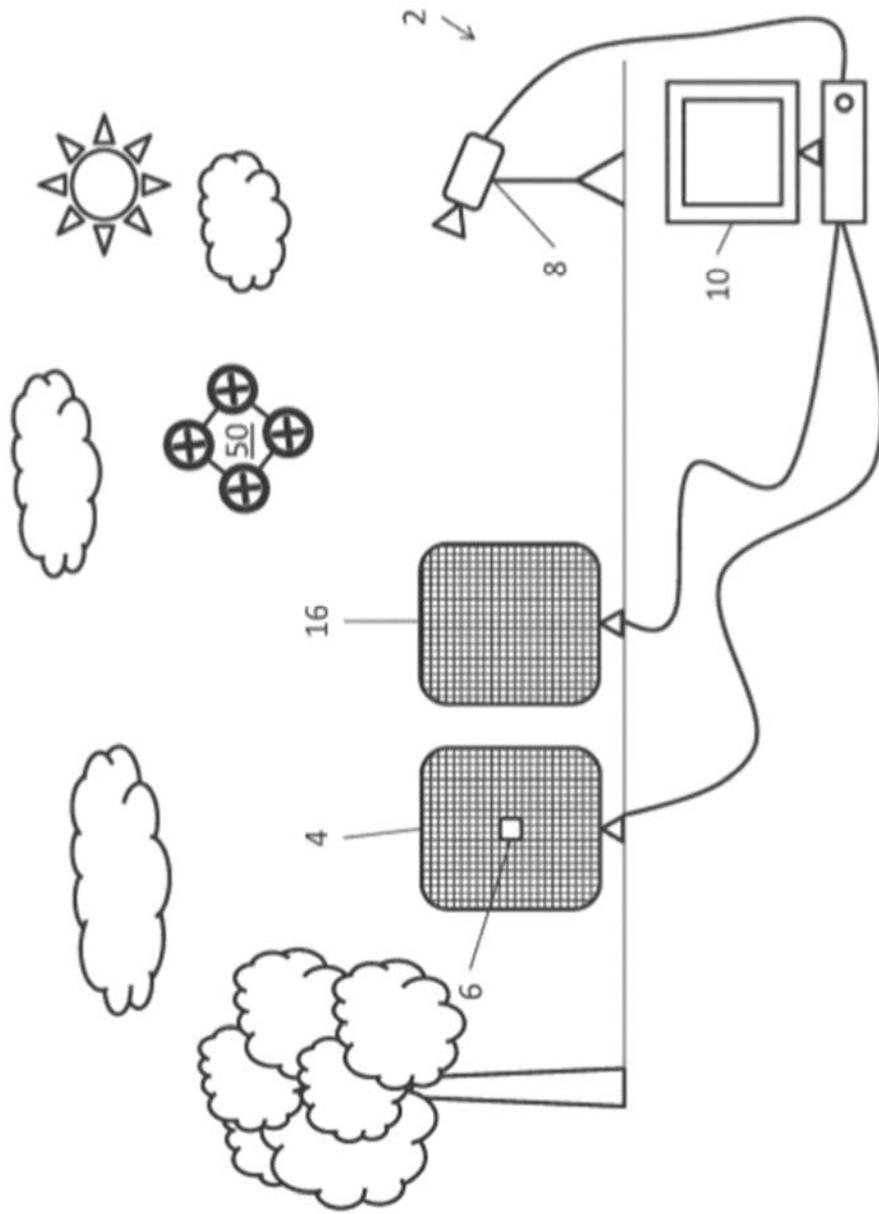


Fig. 9

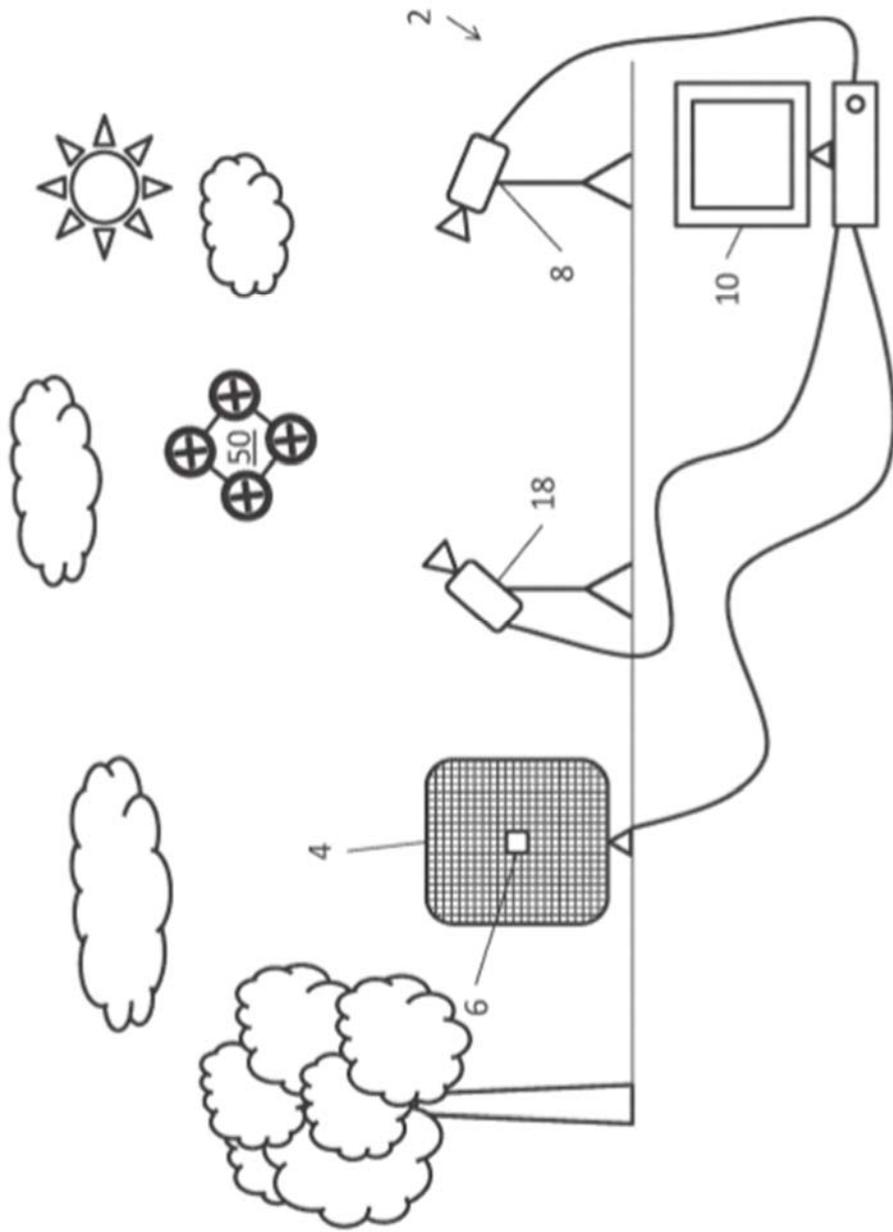


Fig. 10

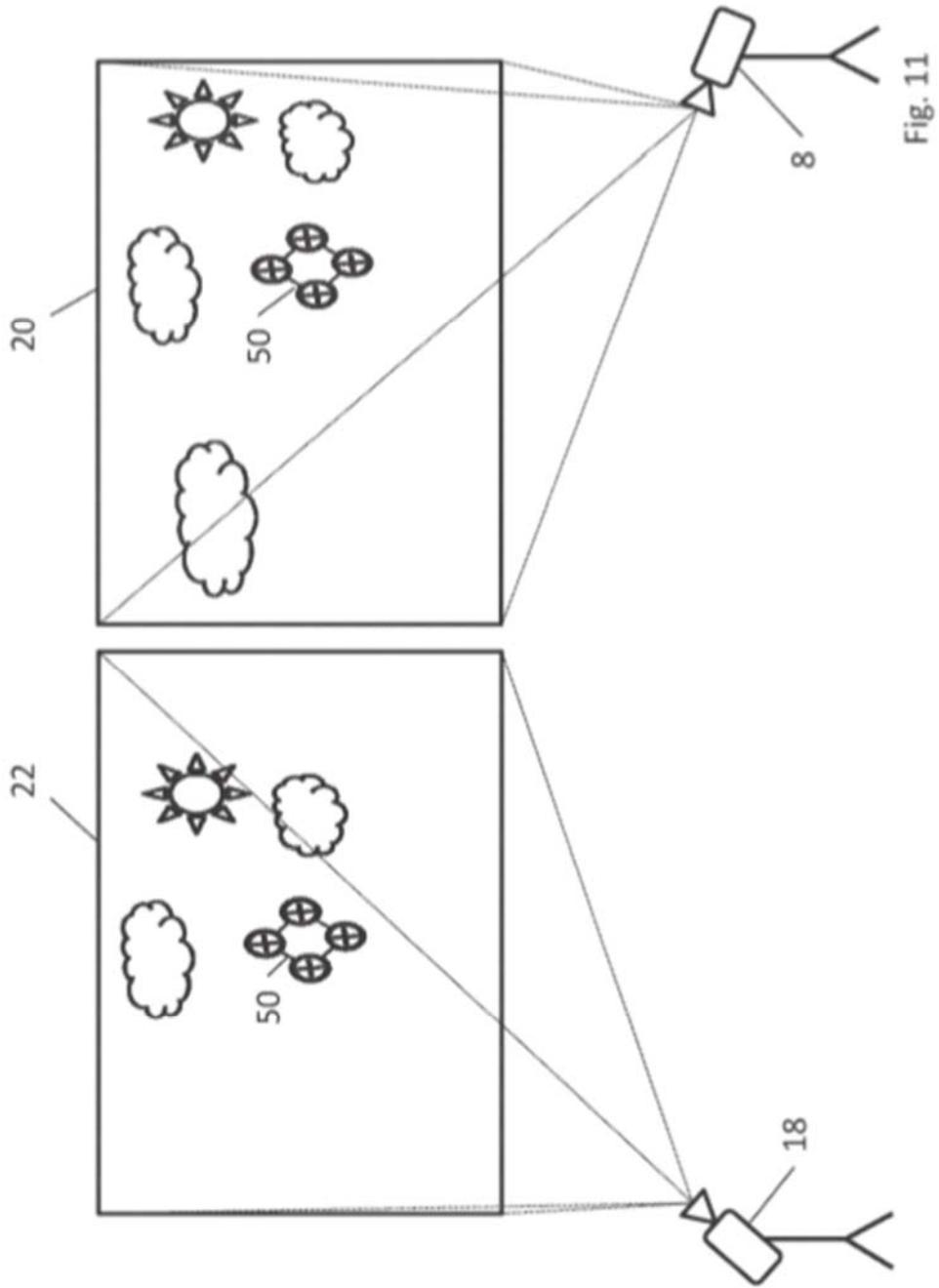


Fig. 11

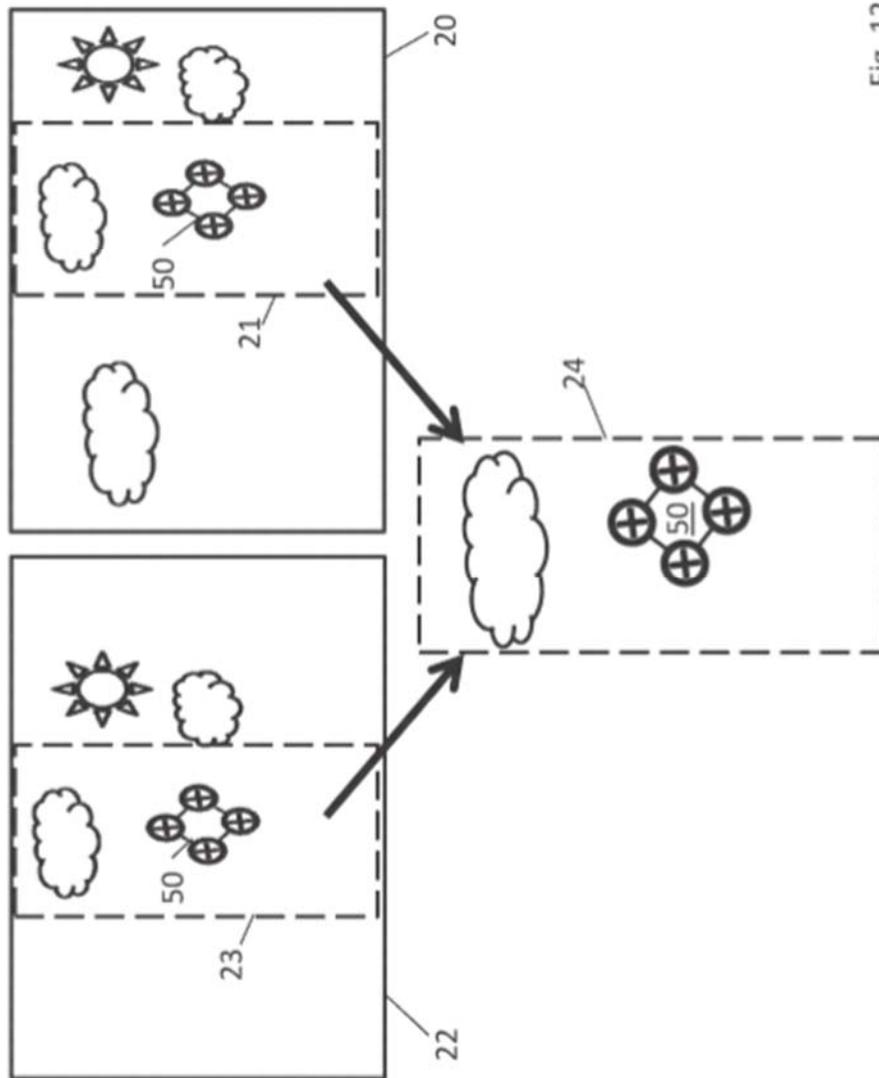


Fig. 12

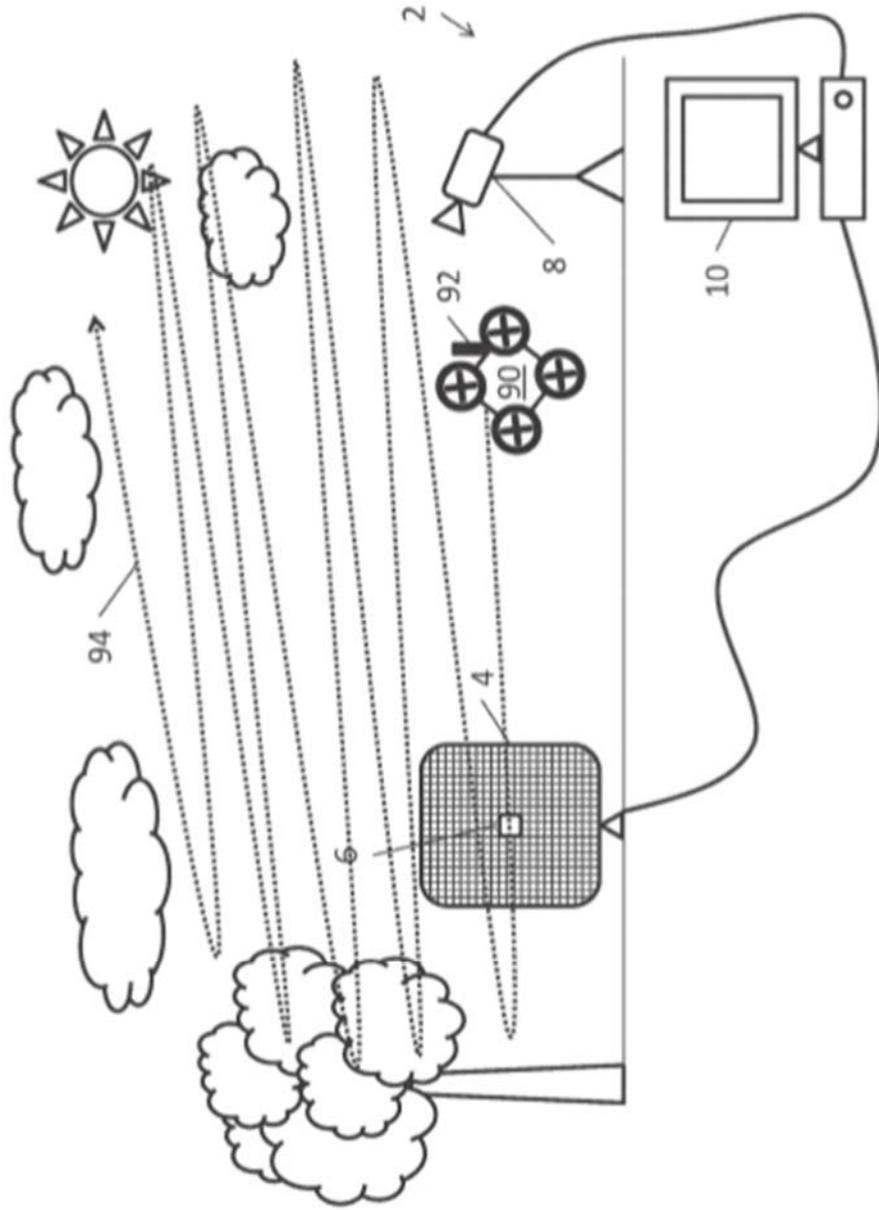


Fig. 13

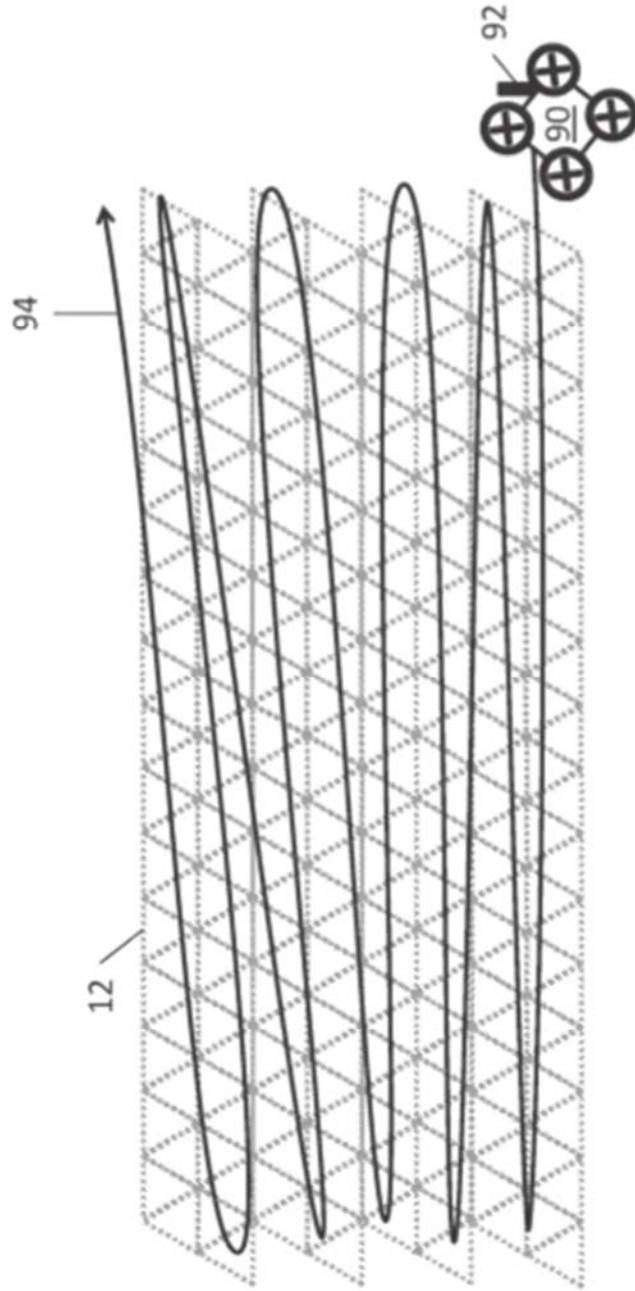


Fig. 14

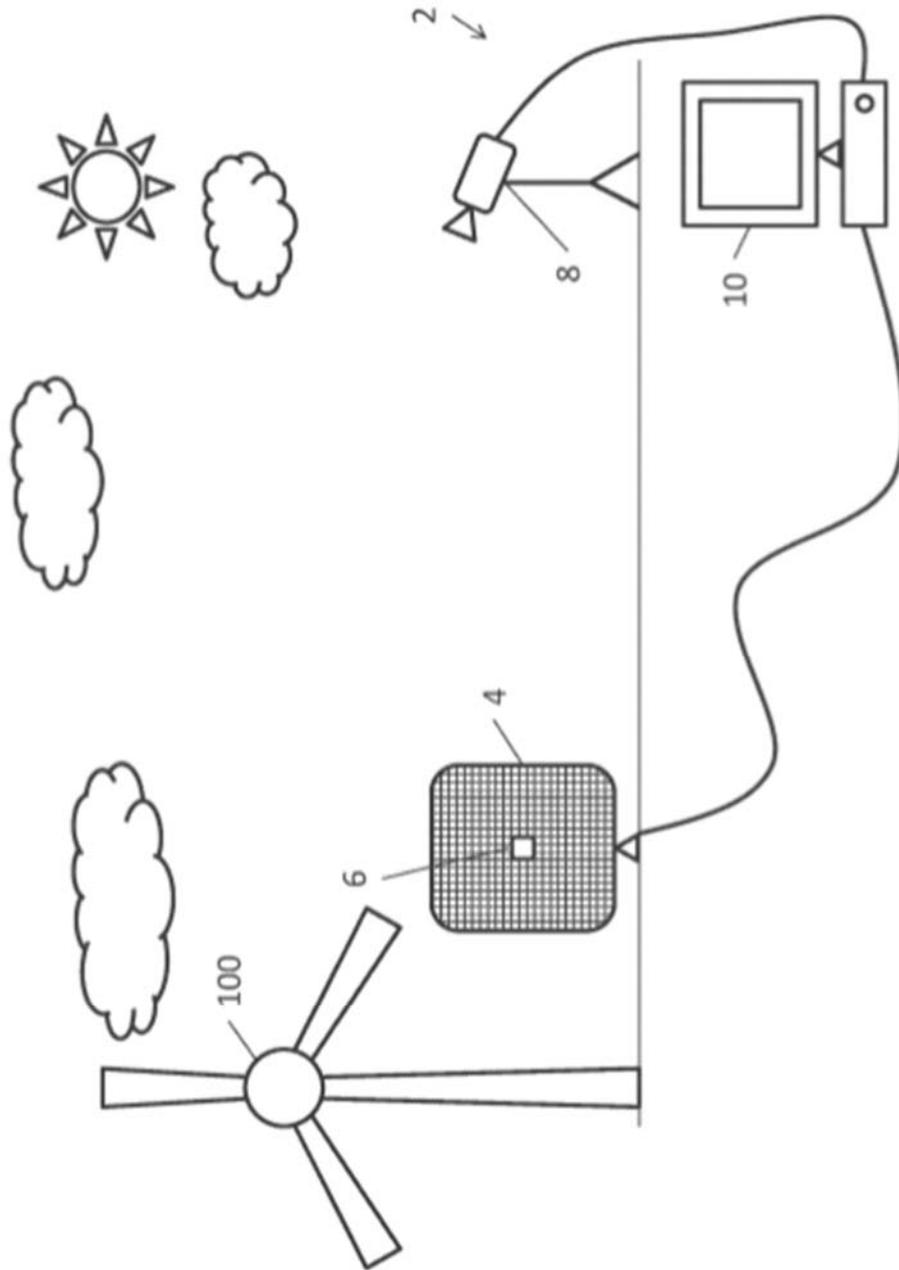


Fig. 15

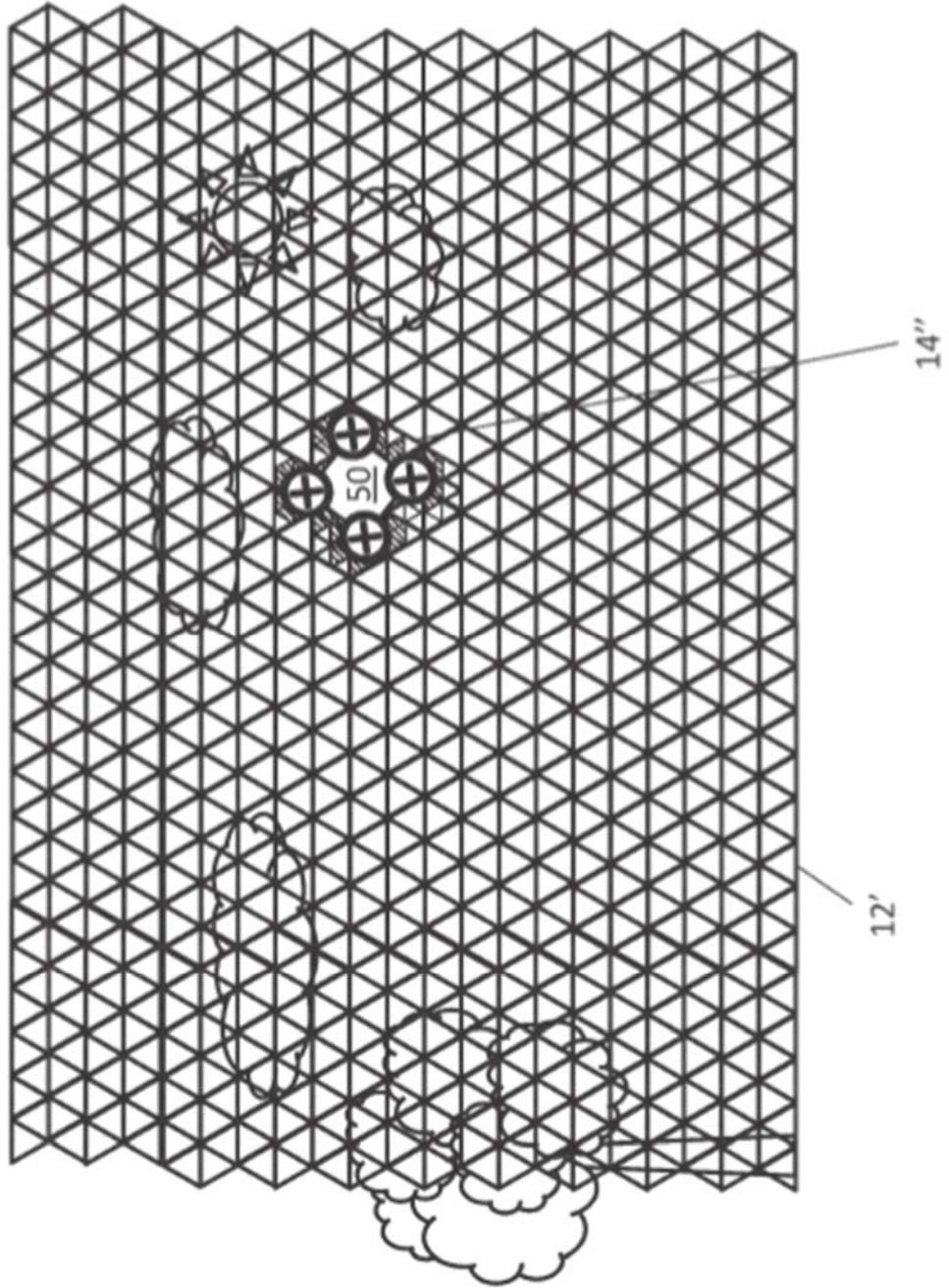


Fig. 16

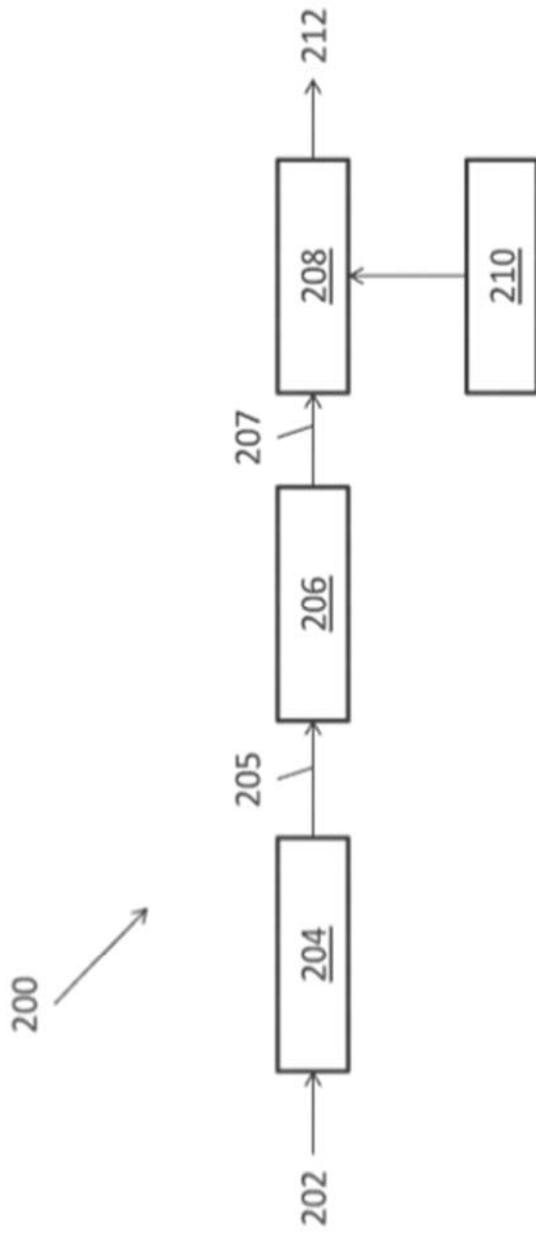


Fig. 17