

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 989 174**

51 Int. Cl.:

G06V 10/24 (2012.01)

G06V 20/56 (2012.01)

G06V 10/82 (2012.01)

G06V 10/774 (2012.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **31.05.2022 PCT/NL2022/050297**

87 Fecha y número de publicación internacional: **08.12.2022 WO22255866**

96 Fecha de presentación y número de la solicitud europea: **31.05.2022 E 22728308 (2)**

97 Fecha y número de publicación de la concesión europea: **24.07.2024 EP 4348595**

54 Título: **Método para entrenar a una red neuronal**

30 Prioridad:

01.06.2021 NL 2028357

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
25.11.2024

73 Titular/es:

**CYCLOMEDIA TECHNOLOGY B.V. (100.0%)
Van Voordenpark 1b
5301 KP Zaltbommel, NL**

72 Inventor/es:

**BOOM, BASTIAAN JOHANNES;
VIJVERBERG, JULIEN ADRIANUS y
BEERS, BART JOHANNES**

74 Agente/Representante:

DEL VALLE VALIENTE, Sonia

ES 2 989 174 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

Método para entrenar a una red neuronal

5 La presente invención se refiere a un método para entrenar a una red neuronal. La presente invención se refiere además a un método para entrenar a una red neuronal para que reconozca un objeto en una imagen, a un método para detectar un objeto en una imagen usando una red neuronal y a un método para detectar un objeto en una imagen.

10 Si se utilizan dos cámaras, por ejemplo, una primera cámara y una segunda cámara, de diferentes fabricantes para grabar una imagen de una misma escena, normalmente se pueden observar distintos colores en las imágenes grabadas. En el caso de las aplicaciones profesionales, este no es siempre un comportamiento deseado. En el caso de los algoritmos de aprendizaje automático, si esos algoritmos aprenden con imágenes captadas por la primera cámara, es probable que el rendimiento de ese algoritmo con imágenes de la segunda cámara sea peor. Una solución actual para abordar este problema es aprender a realizar una transformación de colores entre ambas cámaras. El reto que supone este enfoque es crear conjuntos de datos a gran escala para aprender la transformación requerida.

15 Cabe tener en cuenta que el problema mencionado anteriormente no se limita a las incoherencias en la información de color. Por ejemplo, el uso de cámaras que difieran en al menos una propiedad de entre el grupo de propiedades formado por la resolución máxima de píxeles, la gama de colores, el uso de filtros de color o de polarización, el fabricante de la cámara, el modelo de cámara y los ajustes de exposición, etc., también puede dar lugar al problema de que una red neuronal que haya sido entrenada utilizando imágenes obtenidas con la primera cámara pueda mostrar un rendimiento degradado cuando funcione con imágenes obtenidas con la segunda cámara.

20 En el documento US2018342044 A1 se describe una técnica de mejora de la resolución. Se describe un aparato que recibe unos primeros datos de imagen a una primera resolución y unos segundos datos de imagen a una resolución menor que la primera resolución. Los segundos datos de imagen se pueden ajustar en escala hasta la primera resolución y compararse con los primeros datos de imagen. Una aplicación de una red neuronal puede ajustar en escala los primeros datos de imagen hasta que tengan una resolución mayor que la primera resolución. La aplicación de la red neuronal puede incorporar unas señales basadas en los segundos datos de imagen ajustados a escala. Las señales pueden incluir una información obtenida comparando los segundos datos de imagen ajustados a escala con la resolución de los primeros datos de imagen.

25 En el documento US 2020/372282 A1 se describen un sistema y un método para adaptar imágenes procedentes de distintas cámaras para que se pueda usar un único clasificador o analizador entrenado. El clasificador o analizador funciona con imágenes que incluyen una distribución de colores o unas características de color particular. Se utiliza una red generativa para adaptar unas imágenes procedentes de otras cámaras para que tengan una distribución de colores o una característica de color parecida con el fin de que las pueda utilizar el clasificador o analizador. Para entrenar a la red generativa se utiliza un proceso de confrontación generativa.

30 En el documento KR 102242939 B1 se proporciona un dispositivo de cámara que puede generar imágenes de alta resolución de manera más eficiente. El dispositivo de cámara comprende un sensor de imagen que genera unos primeros datos de Bayer que tienen una primera resolución y un procesador que da salida a unos segundos datos de Bayer que tienen una segunda resolución mayor que la primera resolución mediante la realización de un aprendizaje profundo basado en los primeros datos de Bayer.

35 En el documento US 2017/185851 A1 se describe un sistema que incluye uno o más procesadores y una memoria que incluye unas instrucciones que, cuando son ejecutadas por los uno o más procesadores, hacen que los uno o más procesadores realicen un método. El método incluye recibir unos primeros datos de imagen de un primer sensor de imagen montado en un vehículo y unos segundos datos de imagen de un segundo sensor de imagen montado en el vehículo, incluyendo el vehículo un elemento configurado para abrirse a un espacio externo al vehículo. El método incluye además generar un mapa de profundidad a partir de una comparación de los primeros datos de imagen y los segundos datos de imagen, incluyendo el mapa de profundidad un objeto. Según una determinación de que el objeto, en el mapa de profundidad, está dentro del espacio al que está configurado a abrirse el elemento, se determina que el objeto interferirá con la apertura del elemento al espacio. Además, según una determinación de que el objeto, en el mapa de profundidad, no está dentro del espacio al que está configurado a abrirse el elemento, se determina que el objeto no interferirá con la apertura del elemento al espacio.

40 En el artículo “Vehicular Multi-Camera Sensor System for Automated Visual Inspection of Electric Power Distribution Equipment”, de Park Jinsun y col., Conferencia Internacional del IEEE/RSJ sobre robots y sistemas inteligentes (IROS), IEEE, 3-11-2019, págs. 281-288, se describe un sistema de sensores multicámara para una inspección visual automatizada desde un vehículo en movimiento.

45 La presente invención tiene por objeto proporcionar una manera conveniente de entrenar a una red neuronal en la que los problemas mencionados anteriormente no se produzcan o, al menos, se produzcan en menor medida.

65

Según la presente invención, este objeto se logra usando un método para entrenar a una red neuronal para realizar una transformación de imagen, que comprende proporcionar un vehículo, tal como un coche, en el que se montan una primera cámara y una segunda cámara. La primera cámara y la segunda cámara tienen cada una una misma orientación en relación con el vehículo y están dispuestas espaciadas en una dirección de movimiento predeterminada del vehículo. En el caso de un coche, la dirección de movimiento predeterminada corresponde a una dirección en la que el volante está en una posición no girada.

La primera cámara y la segunda cámara son diferentes la una de la otra con respecto a al menos una propiedad de entre el grupo de propiedades formado por la resolución máxima de píxeles, la gama de colores, el uso de filtros de color o de polarización, el fabricante de la cámara, el modelo de cámara y los ajustes de exposición. En el contexto de la presente invención, se supone que las primera y segunda cámaras tienen idénticas propiedades cuando pertenecen a la misma serie y tienen el mismo número de modelo. Dicho de otra manera, no se tienen en cuenta las variaciones de proceso que son inevitables cuando se fabrican múltiples cámaras que se supone que son idénticas, aunque la invención también podría aplicarse a esas cámaras.

El método según la presente invención comprende además permitir que el vehículo se mueva a lo largo de una trayectoria o controlar el vehículo para que lo haga. Esta trayectoria puede corresponder a una ruta predefinida por una cierta zona. Cuando se graban imágenes para aplicaciones como Street View® de Google®, a menudo se utilizan trayectorias de este tipo.

Según la presente invención, durante el movimiento a lo largo de la trayectoria, se realiza repetidamente una operación de grabación para obtener un conjunto de datos que comprende una pluralidad de pares de imágenes, en donde cada operación de grabación comprende activar la primera cámara y la segunda cámara en distintos instantes en el tiempo, de modo que, en un momento o grabación de una imagen, la pupila de entrada de la primera cámara y la pupila de entrada de la segunda cámara estén sustancialmente en la misma posición a lo largo de la trayectoria. Cada par de imágenes comprende una imagen grabada por la primera cámara y una imagen grabada por la segunda cámara durante una cierta operación de grabación.

Según la presente invención, con el fin de la grabación de una imagen por parte de la primera cámara y de la segunda cámara, de modo que las pupilas de entrada de esas cámaras estén sustancialmente en la misma posición en el momento de grabar las imágenes respectivas, la activación de aquella cámara de entre la primera cámara y la segunda cámara cuya pupila de entrada llegue la última a dicha misma posición se retrasa una cantidad que es igual a la distancia d entre la primera cámara y la segunda cámara en la dirección de movimiento predeterminada dividida por la velocidad v del vehículo en relación con la activación de la otra cámara de entre la primera cámara y la segunda cámara.

Cabe señalar que en algunos casos teóricamente no es posible activar la primera cámara y la segunda cámara de tal manera que, en el momento de grabar una imagen, las pupilas de entrada de esas cámaras estén exactamente en la misma posición en el espacio. Por ejemplo, el vehículo podría realizar un giro o, por lo demás, desviarse de un trayectoria recta. En esos casos, la activación debe realizarse de tal manera que una diferencia entre las posiciones de las pupilas de entrada en los momentos de grabar las imágenes respectivas sea lo más pequeña posible. Esto es particularmente cierto en el caso de una diferencia de posición en una dirección perpendicular al eje óptico de la primera y segunda cámara, ya que esta diferencia está estrechamente relacionada con los fenómenos de paralaje.

Según la presente invención, el conjunto de datos obtenido se usa para entrenar a una red neuronal, tal como una red neuronal convolucional, a que aprenda a realizar una transformación de imagen para transformar una imagen de una escena grabada por la primera cámara en una imagen que imita una imagen de la escena tal como la hubiera grabado la segunda cámara.

Con el método según la invención se pueden formar imágenes de una gran variedad de escenas diferentes. Tener una variedad tan grande mejora mucho la precisión de la red neuronal. La grabación de la gran variedad se hace posible gracias a la combinación del montaje de las cámaras, es decir, ambas cámaras miran en la misma dirección para grabar la misma escena, y la activación particular de estas cámaras, es decir, ambas cámaras graban una imagen desde sustancialmente la misma posición en el espacio a pesar de que el vehículo se esté moviendo.

El método puede comprender además aumentar en escala una resolución de píxeles de unas imágenes grabadas por la primera cámara si estas imágenes tienen una menor resolución de píxeles que las imágenes grabadas por la segunda cámara. Alternativamente, la resolución de píxeles de las imágenes grabadas por la segunda cámara puede reducirse en escala si estas imágenes tienen una mayor resolución de píxeles que las imágenes grabadas por la primera cámara. Tal aumento o reducción en escala se puede realizar antes de usar el conjunto de datos para entrenar a la red neuronal.

El método puede comprender además realizar un registro de imágenes para cada par de imágenes en el conjunto de datos para alinear geoméricamente unos píxeles en la imagen grabada por la primera cámara y unos píxeles en la imagen grabada por la segunda cámara.

Por ejemplo, realizar el registro de imágenes puede comprender, para cada par de imágenes, seleccionar una imagen en el par de imágenes como imagen de origen y seleccionar la otra imagen en el par de imágenes como imagen de destino, determinar una transformación geométrica para asignar la imagen de origen a la imagen de destino usando al menos una de entre una comparación de patrones de intensidad y una comparación de características de imagen y aplicar la transformación geométrica determinada a la imagen de origen para obtener una imagen de origen transformada.

La transformación geométrica se puede determinar comparando la posición de unas características correspondientes en las imágenes. Por ejemplo, en la imagen de origen y en la imagen de destino puede determinarse la posición de una característica, tal como una forma, un color o similar particular. La transformación geométrica debería garantizar que la posición de unas características correspondientes en las primera y segunda imágenes sea sustancialmente idéntica. En vez de realizar una comparación basada en características, se podrían comparar los patrones de intensidad de la primera y segunda imagen.

El método puede comprender además determinar al menos una región en la imagen de origen transformada que comprende unos píxeles para los que no hay ningún píxel correspondiente presente en la imagen de destino y recortar y/o cortar la imagen de origen transformada para excluir la al menos una región determinada. El método puede comprender adicionalmente determinar al menos una región en la imagen de destino que comprende unos píxeles para los que no hay ningún píxel correspondiente presente en la imagen de origen transformada y recortar y/o cortar la imagen de destino para excluir la al menos una región determinada. La imagen de destino y la imagen de origen en el conjunto de datos se pueden sustituir por la imagen de destino recortada/cortada y por la imagen de origen transformada recortada/cortada, respectivamente. En este caso, el método puede comprender además aumentar o reducir en escala la imagen de destino recortada/cortada y la imagen de origen transformada recortada/cortada hasta que tengan una resolución de píxeles de la imagen de origen original o de la imagen de destino original.

Alternativamente, el método puede comprender además determinar al menos una región en la imagen de origen transformada que comprende unos píxeles para los que no hay ningún píxel correspondiente presente en la imagen de destino y construir una máscara binaria para la imagen de origen transformada. La máscara binaria comprende al menos una región que indica unos píxeles en la imagen de origen transformada que tienen unos píxeles correspondientes en la imagen de destino y comprende al menos una región que indica unos píxeles en la imagen de origen transformada que no tienen ningún píxel correspondiente en la imagen de destino. Asimismo, el método puede comprender además determinar al menos una región en la imagen de destino que comprende unos píxeles para los que no hay ningún píxel correspondiente presente en la imagen de origen transformada y construir una máscara binaria para la imagen de destino. La máscara binaria comprende al menos una región que indica unos píxeles en la imagen de destino que tienen unos píxeles correspondientes en la imagen de origen transformada y comprende al menos una región que indica unos píxeles en la imagen de destino que no tienen ningún píxel correspondiente en la imagen de origen transformada.

Además, usar el conjunto de datos obtenido para entrenar a una red neuronal puede comprender usar la máscara binaria de la imagen de destino y la máscara binaria de la imagen de origen transformada para determinar unos píxeles en la imagen de destino y en la imagen de origen transformada que se ignoran cuando se entrena a la red neuronal.

En vez de modificar la imagen de destino y/o la imagen de origen, el entrenamiento de la red neuronal puede comprender determinar en la imagen de origen transformada de cada par de imágenes proporcionadas a la red neuronal unos píxeles para los que existen unos píxeles correspondientes en la imagen de destino de ese par de imágenes y, cuando se entrena a la red neuronal, tener en cuenta solo aquellos píxeles en las imágenes de origen transformada para los que se han encontrado unos píxeles correspondientes en las imágenes de destino correspondientes.

La segunda cámara puede ser una cámara de menor resolución, y la primera cámara, una cámara de mayor resolución, en donde las imágenes grabadas por la primera cámara sufren un problema de sensibilidad a la luz parásita y/o adolecen de halo púrpura. La sensibilidad a la luz parásita (PLS) es un parámetro de rendimiento –particularmente en el caso de los sensores de imagen CMOS– que cuantifica la sensibilidad de sensor a la luz cuando el obturador supuestamente está cerrado. Halo púrpura se refiere a una imagen fantasma desenfocada de color morado o magenta que aparece en una imagen. Ambos efectos son más pronunciados en las cámaras de alta resolución. El método según la presente invención permite mitigar estos efectos al entrenar adecuadamente a la red neuronal a realizar una transformación en las imágenes obtenidas con la primera cámara. Más en particular, el conjunto de datos utilizado para este entrenamiento comprende unas imágenes de alta resolución obtenidas con la primera cámara, imágenes que presentan los efectos mencionados anteriormente, y unas imágenes de baja resolución obtenidas con la segunda cámara, en las que los efectos mencionados anteriormente no están presentes o, al menos, están presentes en menor medida. Luego se entrena a la red neuronal a identificar una transformación de imagen que pueda usarse en las imágenes obtenidas por la primera cámara para eliminar los efectos mencionados anteriormente. En este caso, las imágenes de alta resolución de la primera cámara pueden reducirse en escala antes de usarse en el entrenamiento de la red neuronal.

Se puede utilizar un enfoque similar cuando en la segunda cámara se utilice un filtro de polarización a la hora de grabar imágenes y en la primera cámara no se haga. En este caso, se entrena a la red neuronal a, o bien aplicar un filtro de polarización virtual a las imágenes grabadas por la primera cámara, o bien a eliminar los efectos del filtro de polarización virtual de las imágenes grabadas por la segunda cámara.

5 Según un segundo aspecto, la presente invención proporciona un método para entrenar a una red neuronal adicional para reconocer un objeto, que comprende usar el método que se ha descrito anteriormente para entrenar a una red neuronal a transformar una imagen. El método comprende además proporcionar un conjunto de datos adicional que comprende una pluralidad de imágenes grabadas por la primera cámara y/o por una o más cámaras que son
10 sustancialmente idénticas a la primera cámara, transformar las imágenes del conjunto de datos adicional utilizando la red neuronal entrenada y entrenar a una red neuronal adicional a reconocer objetos en una imagen grabada por la segunda cámara o por una cámara que es sustancialmente idéntica a la segunda cámara usando las imágenes transformadas del conjunto de datos adicional.

15 Con el tiempo, se reúnen grandes conjuntos de datos utilizando la primera cámara y/o una o más cámaras que sean sustancialmente idénticas a la primera cámara. Por ejemplo, durante un cierto período se utilizaron vehículos en los que se había(n) montado la(s) primera(s) cámara(s) para grabar imágenes para diversas aplicaciones. Un ejemplo de una aplicación de este tipo es el reconocimiento de objetos. En esta aplicación, se usa una red neuronal adicional para reconocer un objeto en una imagen. Esta red neuronal adicional se ha entrenado utilizando imágenes grabadas por la
20 primera cámara y/o por una o más cámaras que son sustancialmente idénticas a la primera cámara. La red neuronal adicional mostrará un rendimiento degradado cuando se utilice para reconocer un objeto en una imagen grabada por la segunda cámara o por una cámara que sea sustancialmente idéntica a la segunda cámara. Por otra parte, puede que no sea posible entrenar a la red neuronal adicional utilizando imágenes grabadas por la segunda cámara y/o por una o más cámaras que sean sustancialmente idénticas a la segunda cámara, ya que puede que no haya disponibles
25 suficientes imágenes. Según el segundo aspecto, la presente invención aborda este problema entrenando a una red neuronal a que aprenda a realizar una transformación de imagen para transformar una imagen de una escena grabada por la primera cámara en una imagen que imita una imagen de la escena tal como la habría grabado la segunda cámara y usando esta red neuronal para transformar el conjunto de datos adicional que comprende unas imágenes grabadas por la primera cámara y/o por una o más cámaras que son sustancialmente idénticas a la primera cámara. Este conjunto de datos adicional transformado se puede usar para entrenar a la red neuronal adicional. La red neuronal
30 adicional entrenada de esta manera mostrará un rendimiento mejorado cuando reconozca un objeto en una imagen grabada por la segunda cámara o por una cámara que sea sustancialmente idéntica a la segunda cámara.

35 Según un tercer aspecto, la presente invención proporciona un método para detectar un objeto en una primera imagen, que comprende usar el método descrito anteriormente para entrenar a una red neuronal a transformar una imagen. El método comprende además proporcionar una red neuronal adicional para reconocer un objeto, en donde se ha entrenado a la red neuronal adicional utilizando un conjunto de datos de imágenes grabadas por la segunda cámara y/o por una o más cámaras que son sustancialmente idénticas a las de la segunda cámara. Además, el método comprende grabar una primera imagen usando la primera cámara o una cámara que es sustancialmente idéntica a la
40 primera cámara, transformar la primera imagen usando la red neuronal entrenada y reconocer el objeto en la primera imagen transformada usando la red neuronal adicional entrenada.

La presente invención se describirá a continuación usando los dibujos adjuntos, en donde:

45 la Figura 1 ilustra un vehículo con unas primera y segunda cámaras montadas en el mismo según la presente invención;

La Figura 2 ilustra varios pasos para construir un conjunto de datos que se va a utilizar para entrenar a una red neuronal según la presente invención;

50 la Figura 3 ilustra el entrenamiento de una red neuronal usando el conjunto de datos construido en la Figura 2 según la presente invención;

55 las Figuras 4A y 4B ilustran dos métodos diferentes para reconocer un objeto en una imagen según la presente invención.

La Figura 1 ilustra un vehículo 3 con una primera cámara 1 y una segunda cámara 2 montadas en un techo del vehículo 3. Las cámaras 1, 2 están separadas por una distancia d en una dirección que corresponde a una dirección θ de movimiento predeterminada del vehículo 3. Por ejemplo, el vehículo 3 puede ser un coche, y la dirección θ de movimiento predeterminada corresponde a una dirección de avance en línea recta del vehículo 3.
60

La primera cámara 1 está dirigida a la izquierda del vehículo 3 y tiene un primer campo 4 de visión. Del mismo modo, la segunda cámara 2 también está dirigida a la izquierda del vehículo 3 y tiene un segundo campo 5 de visión que puede o no diferir del campo 4 de visión en un momento de grabarse una imagen.
65

El vehículo 3 se mueve a lo largo de una trayectoria 6 en una carretera 7. A lo largo de esta trayectoria se identifican varias posiciones P_n, P_{n+1}, P_{n+2} en las que la primera cámara 1 y la segunda cámara 2 tienen que grabar una imagen. Para ello, el vehículo 1 puede estar equipado con un sistema de posicionamiento conocido que marca el momento en el que la primera cámara 1 llega a una posición particular, por ejemplo, P_{n+1} . Alternativamente, la primera cámara 1 puede ser activada para grabar una imagen cuando el vehículo 3 haya recorrido una cierta distancia s en relación con una posición en la que la primera cámara 1 ha grabado una imagen anterior. Así pues, es mucho menos importante la precisión de la posición P_{n+1} que garantizar que la primera cámara 1 y la segunda cámara 2 graben una imagen sustancialmente en la misma posición en el espacio. Cuando el vehículo 3 se mueva a una velocidad constante v y la primera cámara 1 haya grabado una imagen en un instante t_0 en el tiempo, la segunda cámara 2 será activada para que grabe una imagen en un instante $t_1 = t_0 + d/v$ en el tiempo.

Gracias al uso de la activación mencionada anteriormente de la primera cámara 1 y de la segunda cámara 2, en combinación con una orientación sustancialmente idéntica de las cámaras 1, 2, se obtienen imágenes de, sustancialmente, la misma escena. Esta escena puede comprender, por ejemplo, un objeto 8.

La primera cámara 1 y la segunda cámara 2 son cámaras diferentes. Más en particular, la primera cámara 1 y la segunda cámara 2 son diferentes entre sí con respecto a al menos una propiedad de entre el grupo de propiedades formado por la resolución máxima de píxeles, la gama de colores, el uso de filtros de color o de polarización, el fabricante de la cámara, el modelo de cámara y los ajustes de exposición.

Con el vehículo 1 se obtiene un conjunto de datos que comprende, para cada una de las M posiciones P_i —donde $i = 1 \dots M$, un par de imágenes que comprenden una imagen grabada por la primera cámara 1 y una imagen grabada por la segunda cámara 2.

La Figura 2 ilustra varios pasos para construir un conjunto de datos que se va a utilizar para entrenar a una red neuronal según la presente invención. Aquí, el conjunto de datos comprende una pluralidad de pares 11 de imágenes. En la Figura 2 se indica un único par 11 que comprende una imagen Im_1 grabada por la primera cámara 1 y una imagen Im_2 grabada por la segunda cámara 2. Tal y como se muestra, la imagen Im_1 tiene una resolución de píxeles más baja que la imagen Im_2 . Además, en cada una de estas imágenes se puede identificar una región 10 que corresponde a un objeto, por ejemplo, el objeto 8, en la Figura 1.

Tal y como se muestra en la fila superior de la Figura 2, la imagen Im_1 y la imagen Im_2 pueden estar geoméricamente desalineadas. Por ejemplo, las imágenes Im_1, Im_2 pueden tener una rotación y/o una traslación mutua(s). Por ejemplo, la posición y la rotación relativas de la región 10 en cada una de las imágenes son diferentes.

Como primer paso, se puede reducir una diferencia en la resolución de píxeles de las imágenes Im_1, Im_2 aumentándolas o reduciéndolas en escala. En la fila central de la Figura 2, la imagen Im_2 se ha reducido en escala utilizando métodos conocidos.

Como paso siguiente, que se muestra en la fila inferior, la imagen Im_1 se somete a una transformación geométrica para alinear geoméricamente las imágenes Im_1, Im_2 . Esta imagen transformada viene indicada por $T(Im_1)$. A pesar de esta alineación geométrica, puede que la posición de la región 10 en estas imágenes no sea exactamente la misma.

Para abordar este problema, se pueden generar unas máscaras binarias que se pueden usar cuando se entrene a la red neuronal. Por ejemplo, se puede generar una máscara binaria $M(Im_2)$ para la imagen Im_2 de destino. Esta máscara binaria tiene una o más regiones $m+$ que comprenden unos píxeles que tienen cada uno un respectivo píxel correspondiente en la imagen $T(Im_1)$ de origen transformada. Además, la máscara binaria $M(Im_2)$ tiene una o más regiones $m-$ que comprenden unos píxeles que no tienen cada uno un respectivo píxel correspondiente en la imagen $T(Im_1)$ de origen transformada. Del mismo modo, se puede generar una máscara binaria $M(T(Im_1))$ para la imagen $T(Im_1)$ de origen transformada. Esta máscara binaria tiene una o más regiones $m+$ que comprenden unos píxeles que tienen cada uno un respectivo píxel correspondiente en la imagen Im_2 de destino. Además, la máscara binaria $M(T(Im_1))$ tiene una o más regiones $m-$ que comprenden unos píxeles que no tienen cada uno un respectivo píxel correspondiente en la imagen $T(Im_2)$ de destino. Durante el entrenamiento de la red neuronal, las máscaras binarias $M(T(Im_1))$ y $M(Im_2)$ se pueden usar para determinar qué píxeles hay que tener o no en cuenta.

En vez de usar máscaras binarias, la imagen $(T(Im_1))$ de origen transformada y la imagen (Im_2) de destino pueden recortarse y/o cortarse para excluir unas regiones en estas imágenes en las que unos píxeles no tienen ningún píxel correspondiente en la otra imagen. En este caso, las imágenes de origen y de destino en el conjunto de datos pueden sustituirse por las homólogas recortadas/cortadas.

La Figura 3 ilustra el entrenamiento de una red 20 neuronal usando un conjunto 12 de datos construido en la Figura 2 según la presente invención. La Figura 3 ilustra un ejemplo en el que el conjunto 12 de datos comprende una pluralidad de pares 11 de imágenes. Cada par 11 comprende una imagen $T(Im_1)$ y una imagen Im_2 , tal como se ha mencionado en relación con la Figura 3.

El conjunto 12 de datos se suministra a una red 20 neuronal para entrenarla. Más en particular, a la red 20 neuronal se la entrena para que las imágenes T(lm1) se transformen para imitar las imágenes lm2. Para ello se utiliza una comparación –indicada aquí como que es ejecutada por un comparador 21– entre las imágenes T(lm1) transformadas por la red 20 neuronal y las imágenes lm2 correspondientes. El resultado de esta comparación se usa para entrenar a la red 20 neuronal. El comparador 21 se puede plasmar usando una función de pérdida para proporcionar una realimentación a la red neuronal.

las Figuras 4A y 4B ilustran dos métodos diferentes para reconocer un objeto en una imagen según la presente invención.

La Figura 4A ilustra un reconocimiento por parte de una red neuronal NN2 de un objeto en una imagen lm2 grabada por la segunda cámara 2 o por una cámara que sea sustancialmente idéntica a la segunda cámara 2. A esta última red neuronal se la entrena, tal y como se ha indicado mediante la línea discontinua, utilizando un conjunto D_0(1)* de datos.

Idealmente, para entrenar a la red neuronal NN2 se utilizan unas imágenes que son parecidas a la imagen lm2. Así, para entrenar a la red neuronal NN2 se usa idealmente un gran conjunto de datos de imágenes grabadas la segunda cámara 2 y/o una o más cámaras que sean sustancialmente idénticas a la segunda cámara 2. La Figura 4A aborda el problema de que no esté disponible tal conjunto de datos. Sin embargo, se encuentra disponible un conjunto D_0(1) de datos que comprende imágenes grabadas por la primera cámara 1 y/o por una o más cámaras que sean sustancialmente idénticas a la primera cámara 2. Por ejemplo, la primera cámara 1 puede corresponder a una cámara conocida y muy utilizada, y la segunda cámara 2 puede corresponder a una cámara novedosa que tenga una mayor resolución, una gama de colores diferente, etc.

Tal y como se muestra en la Figura 4A, la presente invención propone usar el vehículo 1, en el que están montadas una primera cámara 1 y una segunda cámara 2, para grabar un conjunto D_1(1,2) de datos que comprende una pluralidad de pares de imágenes, comprendiendo cada par de imágenes una imagen grabada por la primera cámara 1 y una imagen grabada por la segunda cámara 2. Este conjunto de datos puede modificarse tal y como se muestra en la Figura 3. Además, este conjunto de datos, ya sea en forma original o modificada, se utiliza para entrenar a una red neuronal NN1.

Después de haberse entrenado a la red neuronal NN1, se utiliza para transformar unas imágenes que hay en el conjunto D_0(1) de datos en un conjunto D_0(1)* de datos modificado. Este último conjunto de datos imita un conjunto de datos de imágenes que se habrían obtenido usando la segunda cámara 2. Además, el conjunto D_0(1)* de datos modificado se usa para entrenar a la red neuronal NN2. Como parece que las imágenes del conjunto D_0(1)* de datos modificado y la imagen lm2 se han grabado utilizando la misma cámara o una cámara sustancialmente idéntica, se mejora el rendimiento de la red neuronal NN2 a la hora de reconocer un objeto en una imagen grabada por la segunda cámara 2 o por una cámara que sea sustancialmente idéntica a la segunda cámara 2.

La Figura 4B ilustra un ejemplo en donde se entrena a una red neuronal NN2 a reconocer un objeto en una imagen utilizando un conjunto D_2(2) de datos. Este conjunto de datos comprende unas imágenes que se han grabado usando la segunda cámara 2 y/o una o más cámaras que son sustancialmente idénticas a la segunda cámara 2.

La red neuronal NN2 es menos apta para reconocer un objeto en una imagen grabada por la primera cámara 1. Para abordar este problema, la presente invención propone usar un vehículo 1, en el que están montadas una primera cámara 1 y una segunda cámara 2, para grabar un conjunto D_1(1,2) de datos que comprende una pluralidad de pares de imágenes, comprendiendo cada par de imágenes una imagen grabada por la primera cámara 1 y una imagen grabada por la segunda cámara 2. Este conjunto de datos se utiliza para entrenar a una red neuronal NN1. Esta última red se usa luego para transformar una imagen lm1, grabada por la primera cámara 1 o por una cámara que sea sustancialmente idéntica a la primera cámara 1, en una imagen modificada lm1*. Esta imagen modificada se suministra luego a la red neuronal NN2 para que reconozca un objeto en la misma.

La red neuronal NN1 entrenada se puede utilizar para transformar imágenes que adolezcan de efectos no deseados. Por ejemplo, la primera cámara 1 puede corresponder a una cámara de baja resolución y la segunda cámara 2 puede corresponder a una cámara de alta resolución que, sin embargo, sufra un problema de sensibilidad a la luz parásita y/o adolezca de halo púrpura. En este caso, a la red neuronal NN1 se la entrena para que transforme una imagen introducida obtenida por la segunda cámara 2 en una imagen que imite una imagen que habría sido obtenida por la primera cámara 1. Dicho de otra manera, la red neuronal NN1 está configurada para eliminar los efectos de halo púrpura y/o de sensibilidad a la luz parásita de una imagen obtenida por la segunda cámara 2. En este caso, las imágenes obtenidas por la primera cámara 1 pueden aumentarse en escala hasta que tengan la resolución de píxeles de las imágenes obtenidas por la segunda cámara 2 con el fin de entrenar a la red neuronal NN1.

Cuando se graban imágenes, pueden aparecer problemas por culpa de ventanas que reflejen la luz, de reflexión de la luz solar, etc. Estos problemas se pueden mitigar usando un filtro de polarización. Sin embargo, tener un filtro de polarización instalado en una cámara reduce la entrada de luz, lo cual puede no resultar práctico dados los tiempos de obturación en combinación con el vehículo 3 en movimiento. Este problema puede ser abordado por la presente invención utilizando una segunda cámara 2 y una primera cámara 1 que tenga y no tenga un filtro de polarización,

5 respectivamente. Obsérvese que la segunda cámara 2 y la primera cámara 1 pueden ser idénticas aparte de si se utiliza el filtro de polarización. De este modo, a la red neuronal NN1 se la entrena para transformar una imagen obtenida usando una primera cámara 1 o una cámara que sea sustancialmente idéntica a la primera cámara 1 sin un filtro de polarización en una imagen que imite la imagen que habría obtenido esta cámara si se hubiera utilizado el filtro de polarización. Se puede utilizar un enfoque parecido cuando se utilicen filtros de color.

La presente invención se ha explicado usando unas realizaciones detalladas de la misma. Sin embargo, la presente invención no está limitada a estas realizaciones. Al contrario, son posibles diversas modificaciones sin desviarse del alcance de la presente invención, que está definida por las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un método para entrenar a una red (20) neuronal para realizar una transformación de imagen, que comprende:

5 proporcionar un vehículo, tal como un coche (3), en el que están montadas una primera cámara (1) y una segunda cámara (2), teniendo la primera cámara (1) y la segunda cámara (2) cada una una misma orientación en relación con el vehículo (3) y estando dispuestas espaciadas por una distancia d en una dirección (9) de movimiento predeterminada del vehículo (3), en donde la primera cámara (1) y la segunda cámara (2) son diferentes entre sí con respecto a al menos una propiedad de entre el grupo de propiedades formado por la resolución máxima de píxeles, la gama de colores, el uso de filtros de color o de polarización, el fabricante de la cámara, el modelo de cámara y los ajustes de exposición;

10 permitir que el vehículo (3) se mueva o controlarlo para que se mueva a lo largo de una trayectoria (6);

15 mientras se mueve a lo largo de la trayectoria (6), realizar repetidamente una operación de grabación para obtener un conjunto (11) de datos que comprende una pluralidad de pares (11) de imágenes, en donde cada operación de grabación comprende activar la primera cámara (1) y la segunda cámara (2) en instantes en el tiempo diferentes de modo que en el momento o grabación de una imagen, la pupila de entrada de la primera cámara (1) y la pupila de entrada de la segunda cámara (2) estén sustancialmente en la misma posición (P_n, P_{n+1}, P_{n+2}) a lo largo de la trayectoria (6), en donde cada par (11) de imágenes comprende una imagen (Im1) grabada por la primera cámara (1) y una imagen (Im2) grabada por la segunda cámara (2) durante una cierta operación de grabación;

20 en donde con el fin de la grabación de una imagen por parte de la primera cámara (1) y de la segunda cámara (2) de modo que las pupilas de entrada de esas cámaras estén sustancialmente en la misma posición en el momento de grabar las imágenes respectivas, la activación de aquella cámara de entre la primera cámara (1) y la segunda cámara (2) cuya pupila de entrada llegue la última a dicha misma posición se retrase una cantidad igual a la distancia d entre la primera cámara (1) y la segunda cámara (2) en la dirección (9) de movimiento predeterminada dividida por la velocidad v del vehículo (3) en relación con la activación de la otra cámara de entre la primera cámara (1) y la segunda cámara (2);

25 usar el conjunto de datos obtenido para entrenar a una red (20) neuronal a que aprenda a realizar una transformación de imagen para transformar una imagen de una escena grabada por la primera cámara (1) en una imagen que imite una imagen de la escena tal como la habría grabado la segunda cámara (2).

30

35

2. El método según la reivindicación 1, que comprende además aumentar en escala la resolución de píxeles de unas imágenes (Im1) grabadas por la primera cámara (1) si estas imágenes (Im1) tienen una resolución de píxeles más baja que las imágenes (Im2) grabadas por la segunda cámara (2) o reducir en escala la resolución de píxeles de unas imágenes (Im2) grabadas por la segunda cámara (2) si estas imágenes (Im2) tienen una resolución de píxeles más alta que las imágenes (Im1) grabadas por la primera cámara (1).
- 40

3. El método según la reivindicación 1 o 2, que comprende además realizar un registro de imágenes para cada par (11) de imágenes en el conjunto (12) de datos para alinear geoméricamente unos píxeles en la imagen (Im1) grabada por la primera cámara (1) y unos píxeles en la imagen (Im2) grabada por la segunda cámara (2) antes de usar el conjunto de datos para entrenar a la red (20) neuronal.
- 45

4. El método según la reivindicación 3, en donde la realización del registro de imágenes comprende, para cada par (11) de imágenes:
- 50

seleccionar una imagen (Im1) en el par (11) de imágenes como imagen de origen y seleccionar la otra imagen (Im2) en el par (11) de imágenes como imagen de destino;

determinar una transformación geométrica para asignar la imagen (Im1) de origen a la imagen (Im2) de destino usando al menos una de una comparación de patrones de intensidad y una comparación de características de imagen;

55 aplicar la transformación geométrica determinada a la imagen (Im1) de origen para obtener una imagen ($T(Im1)$) de origen transformada.

5. El método según la reivindicación 4, que comprende, además:
- 60

determinar al menos una región en la imagen ($T(Im1)$) de origen transformada que comprende unos píxeles para los que no hay ningún píxel correspondiente presente en la imagen (Im2) de destino y recortar y/o cortar la imagen ($T(Im1)$) de origen transformada para excluir la al menos una región determinada;

- determinar al menos una región en la imagen (Im2) de destino que comprende unos píxeles para los que no hay ningún píxel correspondiente presente en la imagen (T(Im1)) de origen transformada, y recortar y/o cortar la imagen (Im2) de destino para excluir la al menos una región determinada; y sustituir la imagen (Im2) de destino por la imagen de destino recortada/cortada y sustituir la imagen (Im2) de origen por la imagen (T(Im1)) de origen transformada recortada/cortada.
- 5
6. El método según la reivindicación 5, que comprende además aumentar o reducir en escala la imagen (Im2) de destino recortada/cortada y la imagen (T(Im1)) de origen transformada recortada/cortada hasta que tengan una resolución de píxeles en la imagen (Im1) de origen original o en la imagen (Im2) de destino original.
- 10
7. El método según la reivindicación 4, que comprende, además:
- determinar al menos una región en la imagen (T(Im1)) de origen transformada que comprende unos píxeles para los que no hay ningún píxel correspondiente presente en la imagen (Im2) de destino y construir una máscara binaria (M(T(Im1))) para la imagen (T(Im1)) de origen transformada, comprendiendo dicha máscara binaria (M(T(Im1))) al menos una región (m+) que indica unos píxeles en la imagen (T(Im1)) de origen transformada que tienen unos píxeles correspondientes en la imagen (Im2) de destino, y comprendiendo al menos una región (m-) que indica unos píxeles en la imagen (T(Im1)) de origen transformada que no tienen ningunos píxeles correspondientes en la imagen (Im2) de destino;
- 15
- determinar al menos una región en la imagen (Im2) de destino que comprende unos píxeles para los que no hay ningunos píxeles correspondientes en la imagen (T(Im1)) de origen transformada, y construir una máscara binaria (M(Im2)) para la imagen (Im2) de destino, comprendiendo dicha máscara binaria (M(Im2)) al menos una región (m+) que indica unos píxeles en la imagen (Im2) de destino que tienen unos píxeles correspondientes en la imagen (T(Im1)) de origen transformada, y comprendiendo al menos una región (m-) que indica unos píxeles en la imagen (Im2) de destino que no tienen ningunos píxeles correspondientes en la imagen (T(Im1)) de origen transformada.
- 20
- 25
8. El método según la reivindicación 7, en donde dicho uso del conjunto de datos obtenido para entrenar a una red (20) neuronal comprende usar la máscara binaria (M(Im2)) de la imagen (Im2) de destino y la máscara binaria (M(T(Im1))) de la imagen (T(Im1)) de origen transformada para determinar unos píxeles en la imagen (Im2) de destino y en la imagen (T(Im1)) de origen transformada que se ignoran cuando se entrena a la red (20) neuronal.
- 30
- 35
9. El método según la reivindicación 4, en donde el entrenamiento de la red (20) neuronal comprende:
- determinar en la imagen (T(Im1)) de origen transformada de cada par (11) de imágenes proporcionadas a la red (20) neuronal unos píxeles para los que existen unos píxeles correspondientes en la imagen (Im2) de destino de ese par (11) de imágenes; y
- 40
- cuando se entrena a la red (20) neuronal, tener en cuenta únicamente aquellos píxeles en las imágenes (T(Im1)) de origen transformadas para los que se han encontrado unos píxeles correspondientes en las imágenes (Im2) de destino correspondientes.
- 45
10. El método según cualquiera de las reivindicaciones anteriores, en donde la segunda cámara (2) es una cámara de menor resolución y la primera cámara (1) es una cámara de mayor resolución, en donde unas imágenes grabadas por la primera cámara (1) sufren un problema de sensibilidad a la luz parásita y/o adolecen de halo púrpura.
- 50
11. El método según cualquiera de las reivindicaciones anteriores, en donde en la segunda cámara (2) se utiliza un filtro de polarización a la hora de grabar imágenes.
- 55
12. El método según cualquiera de las reivindicaciones anteriores, que además comprende:
- grabar una primera imagen usando la primera cámara (1) o una cámara que es sustancialmente idéntica a la primera cámara y transformar la primera imagen usando la red (20) neuronal entrenada.
- 60
13. El método para entrenar a una red neuronal (NN2) adicional para reconocer un objeto, que comprende:
- usar el método de cualquiera de las reivindicaciones anteriores para entrenar a una red neuronal (NN1) a transformar una imagen;
- 65
- proporcionar un conjunto (D_0(1)) de datos adicional que comprende una pluralidad de imágenes grabadas por la primera cámara (1) y/o por una o más cámaras que son sustancialmente idénticas a la primera cámara (1);
- transformar las imágenes del conjunto (D_0(1)) de datos adicional utilizando la red neuronal (NN1) entrenada;

entrenar a una red neuronal (NN2) adicional para reconocer unos objetos en una imagen (Im2) grabada por la segunda cámara (2) o por una cámara que es sustancialmente idéntica a la segunda cámara (2) usando las imágenes transformadas del conjunto $D_0(1)^*$ de datos adicional.

5 14. Un método para detectar un objeto en una segunda imagen usando una red neuronal (NN2) adicional, que comprende:

entrenar a la red neuronal (NN2) adicional según la reivindicación 13;
grabar una segunda imagen (Im2) por parte de la segunda cámara (2) o de una cámara que es sustancialmente idéntica a la segunda cámara (2);
10 reconocer el objeto en la segunda imagen (Im2) utilizando la red neuronal (NN2) adicional entrenada.

15. Un método para detectar un objeto en una primera imagen (Im1), que comprende:

15 usar el método de cualquiera de las reivindicaciones 1-12 para entrenar a una red neuronal (NN1) a transformar una imagen;
proporcionar una red neuronal (NN2) adicional para reconocer un objeto que ha sido entrenada utilizando un conjunto ($D_2(2)$) de datos de unas imágenes grabadas por la segunda cámara (2) y/o por una o más cámaras que son sustancialmente idénticas a la segunda cámara (2);
grabar una primera imagen (Im1) usando la primera cámara (1);
20 transformar la primera imagen (Im1) usando la red neuronal (NN1) entrenada;
reconocer el objeto en la primera imagen transformada (Im1*) utilizando la red neuronal (NN2) adicional entrenada.

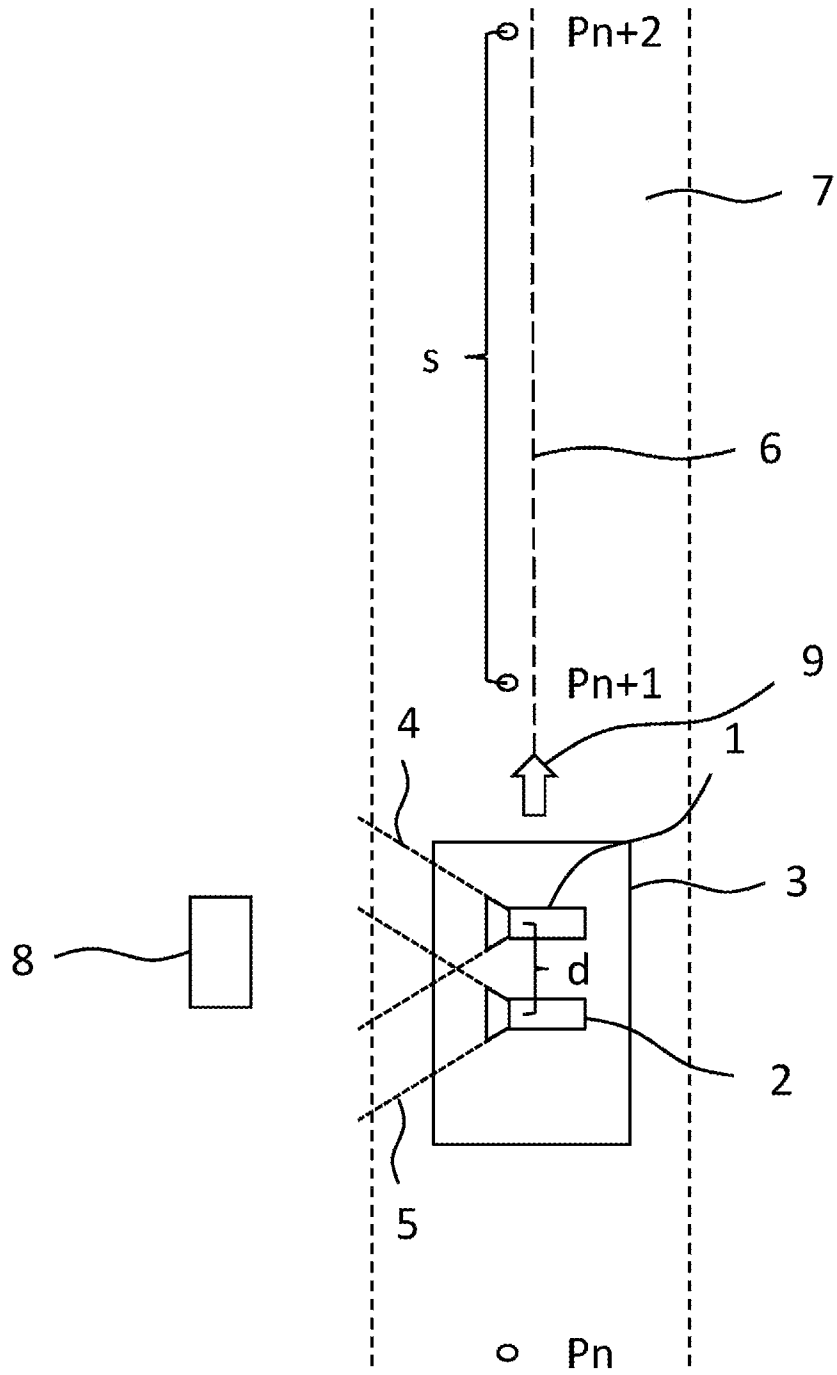


Figura 1

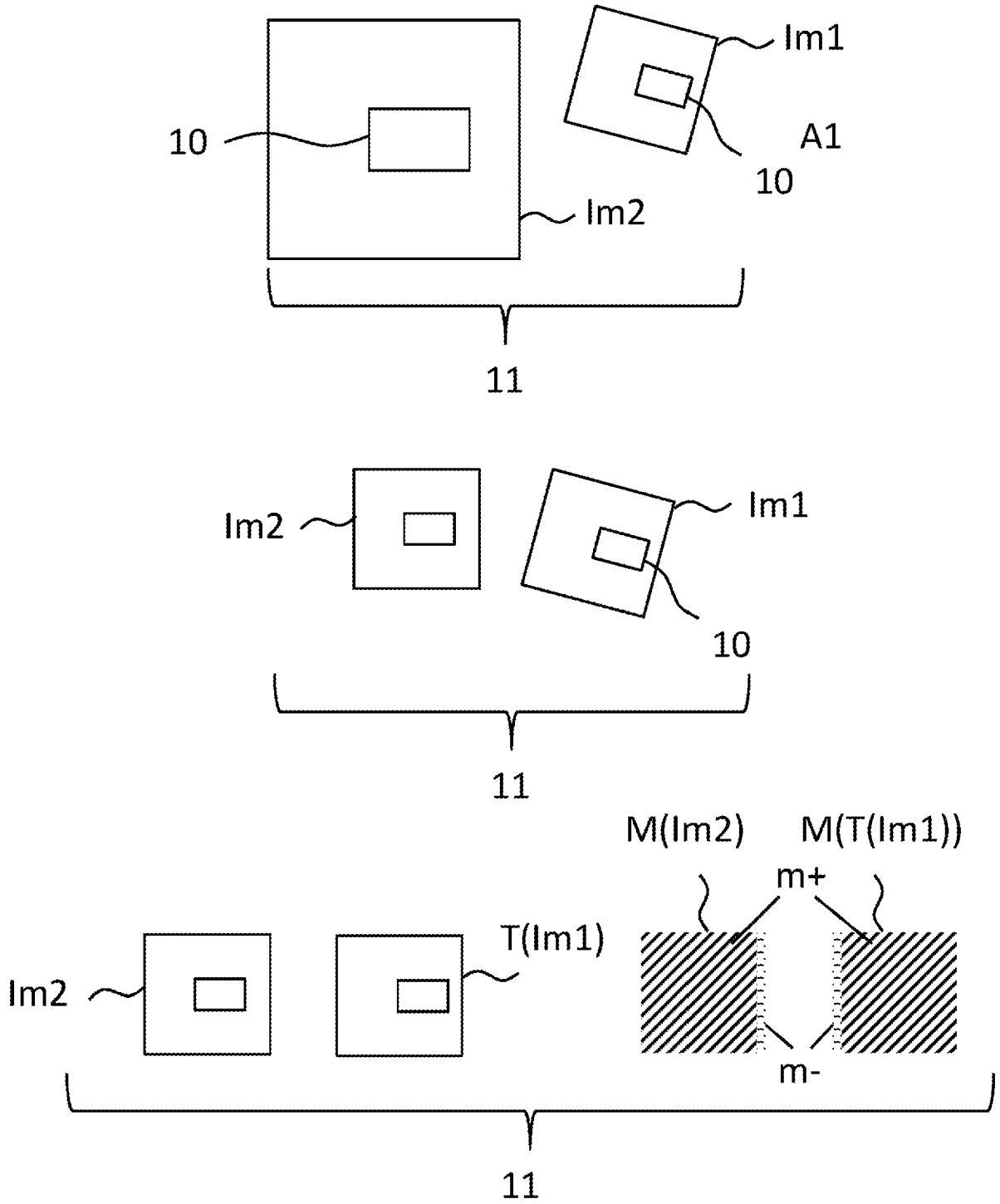


Figura 2

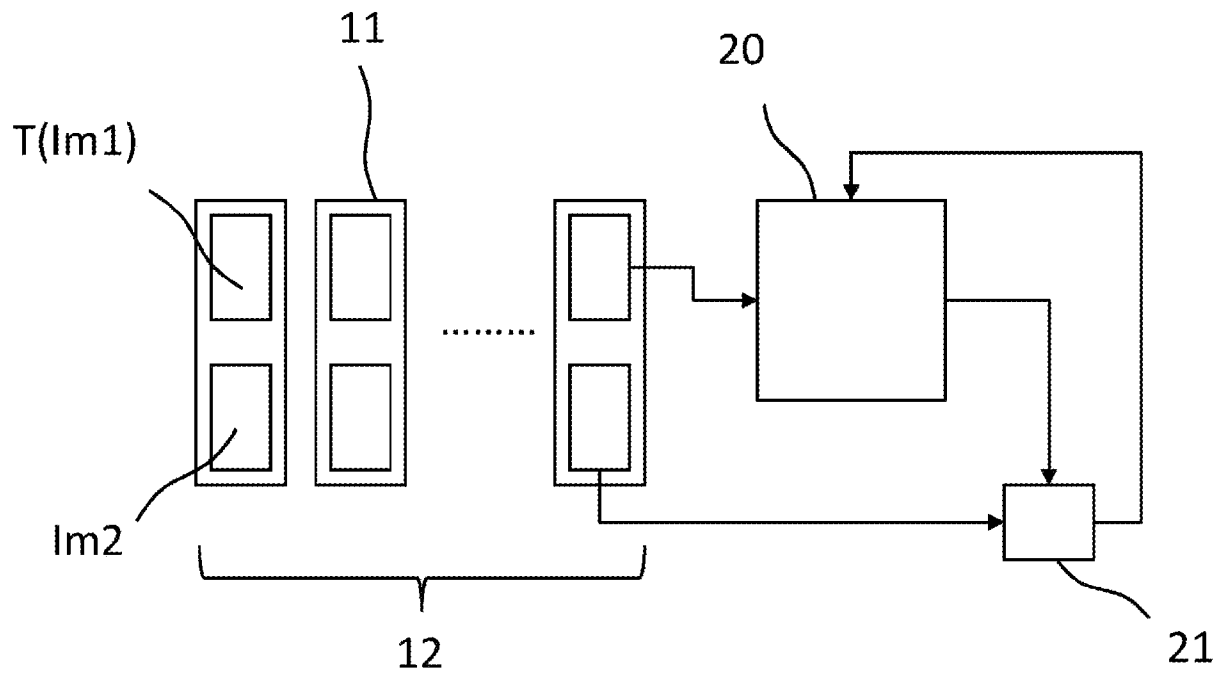


Figura 3

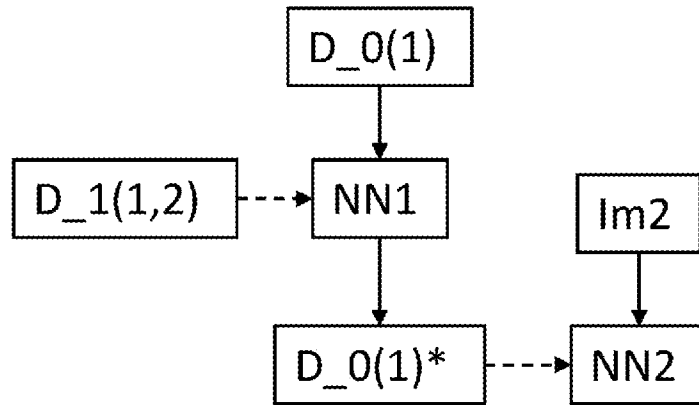


Figura 4A

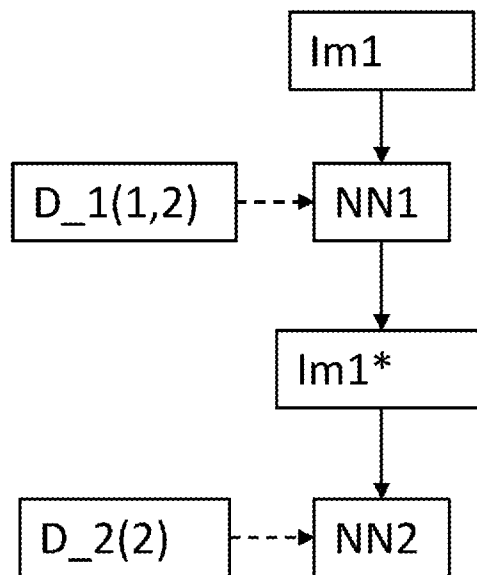


Figura 4B