VNiVERSiDAD D SALAMANCA

CAMPUS DE EXCELENCIA INTERNACIONAL

DEPARTAMENTO DE INGENIERÍA CARTOGRÁFICA

Y DEL TERRENO

Tesis Doctoral

ESTUDIO Y ANÁLISIS DE SENSORES RGB-D DE BAJO COSTE O "GAMING SENSORS" EN APLICACIONES DE VISIÓN ARTIFICIAL

Programa de Doctorado:

Geotecnologías aplicadas a la Construcción, Energía e Industria

Manuel Gesto Díaz

2018

Departamento de Ingeniería Cartográfica y del Terreno

Escuela Politécnica Superior de Ávila

Universidad de Salamanca

Autor:

Manuel Gesto Díaz

Directores:

Dr. Diego González Aguilera

PD Dr. Ing. Habil. Federico Tombari

2018

"ESTUDIO Y ANÁLISIS DE SENSORES RGB-D DE BAJO COSTE O "GAMING SENSORS" EN APLICACIONES DE VISIÓN ARTIFICIAL"

presentada por D. Manuel Gesto Díaz

Informe de los directores

La Tesis Doctoral presentada por el doctorando D. Manuel Gesto Díaz reúne todos los requisitos que se le pueden exigir a una Tesis Doctoral, y que a continuación se pasan a detallar:

El tema presentado por el doctorando en su Tesis Doctoral se enmarca en el campo de la sensórica, la reconstrucción 3D, la metrología, el reconocimiento y la correspondencia de imágenes, alineándose a la perfección con el Programa de Doctorado.

La metodología y algoritmos desarrollados en la Tesis Doctoral presentan aspectos novedosos en el campo de la visión artificial que permiten abordar con éxito los objetivos propuestos.

Por otro lado, hay que reseñar la posible transferencia de tecnología derivada en forma del desarrollo de softwares y patentes cuya propiedad intelectual ha sido registrada en la Universidad de Salamanca y en el Ministerio de Industria, respectivamente y de las que el doctorando es coautor y coinventor.

Finalmente, hay que destacar muy especialmente el nivel de producción científica derivado del propio desarrollo de la Tesis Doctoral por parte del Doctorando, el cual permite avalar la calidad y relevancia de la misma. Hay que reseñar la publicación de 3 artículos indexados JCR asociados a la Tesis Doctoral, uno de ellos en la primera revista dentro de la categoría de "Remote Sensing".

Dada la participación de coautores en los artículos presentados, se relaciona a continuación la aportación de cada autor, en orden cronológico de publicación:

Artículo 1: *"Metrological comparison between Kinect I and Kinect II sensors"*

Higinio González Jorge: coordinación de autores, diseño de experimento y procesamiento de resultados.

Pablo Rodríguez Gonzálvez: asesoramiento y formación en aspectos estadísticos, procesamiento de resultados estadísticos.

Joaquín Martínez Sánchez: elección de sensores para el estudio, calibración de los mismos.

Diego González Aguilera: asesoramiento en la realización de tareas de procesado, así como de explotación de resultados.

Pedro Arias Sánchez: comparación de datos en la nube de puntos para obtener los resultados estadisticos.

Manuel Gesto Díaz: desarrollo de software de adquisición de datos, procesamiento de la nube para su uso en la comparación.

Lucía Díaz Vilariño: realización del estado del arte y redacción del artículo.

Artículo 2: “*Analysis and evaluation between the first and the second generation of RGB-D sensors”*

Manuel Gesto Díaz: diseño de software de adquisición, procesamiento de datos, metodología y valoración de resultados.

Federico Tombari: formación y tutorización de los trabajos desarrollados por el doctorando, supervisión y asesoramiento en la interpretación de resultados.

Pablo Rodríguez Gonzálvez: asesoramiento y formación en aspectos estadísticos, procesamiento de resultados estadísticos.

Diego González Aguilera: asesoramiento en la realización de tareas de procesado, así como de explotación de resultados.

Artículo 3: “*Feature matching evaluation for multimodal correspondence”*

Manuel Gesto Díaz: diseño de software de adquisición, procesamiento de datos, metodología y valoración de resultados.

Federico Tombari: formación y tutorización de los trabajos desarrollados por el doctorando, supervisión y asesoramiento en la interpretación de resultados.

Diego González Aguilera: asesoramiento en la realización de tareas de procesado, así como de explotación de resultados.

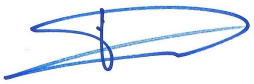
Luis López Fernández: programación de diferentes algoritmos de visión artificial utilizados en la comparación

Pablo Rodríguez Gonzálvez: asesoramiento y formación en aspectos estadísticos, procesamiento de resultados estadísticos.

Por todo lo anteriormente reseñado, emito un informe con todos mis pronunciamientos favorables, y autorizo su presentación como Tesis Doctoral en el Departamento de Ingeniería Cartográfica y del Terreno de la Universidad de Salamanca.

Ávila, 15 de agosto de 2018

LOS DIRECTORES DE LA TESIS DOCTORAL



Fdo. Diego González Aguilera

Fdo. Federico Tombari

Listado de artículos publicados

La presente Tesis Doctoral está constituida por un compendio de tres artículos científicos, publicados en revistas internacionales de alto impacto. A continuación, se enumeran estas publicaciones:

1. Metrological comparison between Kinect I and Kinect II sensors

H. Gonzalez-Jorge [a,] P. Rodríguez-Gonzálvez [b], J. Martínez-Sánchez [a], D. González-Aguilera [b],P. Arias a, M. Gesto [b], L. Díaz-Vilariño [a]

[a] Applied Geotechnology Group, School of Mining Engineering, University of Vigo, Rúa Maxwell s/n, Campus Lagoas-Marcosende, 36310 Vigo, Spain

[b] TIDOP Research Group, Cartographic and Land Engineering Department, High Polytechnic School of Avila, University of Salamanca, 05003 Avila, Spain

Measurement, Marzo 2015

DOI: 10.1016/j.measurement.2015.03.042

2. Analysis and evaluation between the first and the second generation of RGB-D sensors

M. Gesto-Diaz [a], F. Tombari [b,c], P. Rodriguez-Gonzalvez [a] , D. Gonzalez-Aguilera[a]

[a] Cartographic and Land Engineering Department, University of Salamanca, Hornos Caleros 50, 05003 Avila, Spain

[b] DISI, University of Bologna, V.le del Risorgimento 2, Bologna, Italy

[c] CAMP, Technische Universität München (TUM), Boltzmannstr. 3, Garching b. München, Germany

IEEE Sensors Journal , Julio 2015

DOI: 10.1109/JSEN.2015.2459139

3. Feature matching evaluation for multimodal correspondence

M. Gesto-Diaz [a], F. Tombari [b,c], D. Gonzalez-Aguilera [a], L. Lopez-Fernandez [a], P. Rodriguez-Gonzalvez [a]

[a] Cartographic and Land Engineering Department, University of Salamanca, Hornos Caleros 50, 05003 Avila, Spain

[b] DISI, University of Bologna, V.le del Risorgimento 2, Bologna, Italy

[c] CAMP, Technische Universität München (TUM), Boltzmannstr. 3, Garching b. München, Germany

ISPRS Journal of Photogrammetry and Remote Sensing, Mayo 2017

DOI: 10.1016/j.isprsjprs.2017.05.007

A todos los que han sido parte de mi vida y han forjado mi carácter.

"La ciencia(vida) se compone de errores, que, a su vez, son los pasos hacia la verdad"

Julio Verne- escritor, poeta y dramaturgo francés

"Estando siempre dispuestos a ser felices, es inevitable no serlo alguna vez."

Blaise Pascal- polímata, matemático, físico, filósofo y escritor francés

Resumen

Los sensores RGB-D de bajo coste o “gaming sensors” capaces de capturar una imagen con profundidad han supuesto un punto de inflexión en el campo de la visión artificial, haciendo estos dispositivos más atractivos para la comunidad científica, ya que con tan bajo coste se abre la posibilidad de su uso en innumerables aplicaciones de visión artificial y robótica. Desde la irrupción de la primera generación de este dispositivo el número de publicaciones científicas donde se ha utilizado ha crecido exponencialmente. Propiciado por el éxito de este primer dispositivo, se ha lanzado una segunda versión, que, aunque presenta similares características, la tecnología de adquisición de profundidad es diferente, creando el interés en realizar estudios para su comparación y propiciando la idea para la realización de esta tesis.

En la presente tesis doctoral se realizará un análisis de las dos versiones de este dispositivo, primeramente, desde un punto de vista puramente teórico, y a continuación, desde un punto de vista de rendimiento en aplicaciones de visión artificial. La finalidad de estas pruebas es tener los suficientes datos o pruebas que ayuden a identificar que dispositivo o generación de dispositivos se adapta mejor a la aplicación deseada.

El primer análisis es teórico, se evalúa desde un punto de vista metrológico el comportamiento de las dos generaciones de sensores. Conocido el comportamiento teórico se procede a realizar un análisis de rendimiento en diferentes aplicaciones de visión artificial. Se centra este rendimiento en unas aplicaciones específicas, siendo estas, la reconstrucción 3D, el registro de imágenes, el reconocimiento de objetos y finalmente se busca la combinación con múltiples sensores a través del registro de imágenes multimodales. Durante la realización de los experimentos se han logrado numerosos hitos de la mano de la redacción de esta tesis. Entre ellos se encuentra la creación de diferentes bancos de prueba o “datasets”, software y patentes.

Con estos resultados se provee a la comunidad científica de información suficiente acerca del comportamiento de los sensores y de su posibilidad de uso en diferentes aplicaciones.

Abstract

The low cost RGB-D sensors or "gaming sensors" capable of capturing an image with depth have supposed a turning point in the field of artificial vision, making these devices more attractive for the scientific community, since with such a low cost opens the possibility of its use in countless applications of artificial vision and robotics. Since the irruption of the first generation of this device the number of scientific publications where it has been used has grown exponentially. Propelled by the success of this first device, a second version has been launched, which, although it has similar characteristics, the depth acquisition technology is different, creating interest in carrying out studies for comparison and promoting the idea for the realization of this thesis.

In the present doctoral thesis an analysis of the two versions of this device will be carried out, first, from a purely theoretical point of view, and then, from a performance point of view in artificial vision applications. The purpose of these tests is to have enough data or evidence to help identify which device or device generation is best suited to the desired application.

The first analysis is theoretical, the behaviour of the two sensor generations is evaluated from a metrological point of view. Once the theoretical behaviour is known, a performance analysis is carried out in different artificial vision applications. This performance is focused on specific applications, such as 3D reconstruction, image registration, object recognition and finally the combination with multiple sensors is sought through multimodal image registration. During the realization of the experiments numerous milestones have been achieved by the writing of this thesis. Among them is the creation of different test banks or "datasets", software and patents.

With these results, the scientific community is provided with enough information about the behaviour of the sensors and their possibility of use in different applications.

Agradecimientos

Una vez finalizado el gran esfuerzo que conlleva la escritura de una tesis, comenzada hace ya cinco años, es inevitable acordarte de todas las personas e instituciones que te han ayudado a conseguir esta meta y agradecerles, no solo por la consecución del objetivo de escribir una tesis, con las complicaciones que esto conlleva, sino también agradecer el camino que hemos vivido juntos y que me han aportado unas capacidades, aptitudes y madurez que continuarán durante el resto de mi viaje por esta vida.

En primer lugar, quiero dar las gracias a todos los compañeros, personal decente y directores de la tesis, que con su encomendable ayuda han logrado, junto conmigo, la escritura de esta tesis. En especial quiero agradecer a mis directores de la tesis el **Catedrático Diego González Aguilera** y el **PD Dr. Ing. Habil. Federico Tombari**. Primero agradecer la confianza depositada en mi para la consecución de la tesis, apoyando mi propuesta investigadora y formándome para conseguir su desarrollo. Ellos han sido una gran ayuda durante toda la tesis, siempre les estaré agradecido por la gran ayuda que me han prestado y por todas las enseñanzas que me han transmitido, que siempre han ido más allá de lo exigido académicamente, demostrando su gran interés por la investigación y su calidad como personas. De ellos he aprendido a amar la ciencia, la búsqueda por indagar en lo desconocido y enseñar como difundir ese conocimiento entre la comunidad científica. Mi agradecimiento también al Doctor Pablo Rodríguez por el asesoramiento y asistencia que me ha brindado durante este periodo y su inestimable ayuda.

Me gustaría también dar las gracias a mis compañeros del grupo **Tidop**. **Alberto, Susana, Mónica, Sandra, Luis Javier** y especialmente a **Luis López**. Quiero agradecerles a todas las horas compartidas, vivencias, aventuras y por este camino del conocimiento al que todos nos hemos enfrentado hasta la consecución final de esta tesis. Dar las gracias también al grupo de investigación **Applied Geotechnologies** de la Universidad de Vigo por la ayuda prestada.

Finalmente, en el ámbito personal quiero dar las gracias a **Marta**, has sido un gran apoyo durante el proceso de realizar esta tesis. Sólo darte las gracias y espero que sigas siendo mi compañera en esta montaña rusa que es la vida.

Dar las gracias a todos mis amigos, la familia que se escoge, y que tantas veces me habéis ayudado todos estos años.

Y como no, a toda mi **familia**, no ha sido fácil, un difícil camino, pero me habéis enseñado que siempre hay salida, que hay que seguir hacia adelante, por muy difícil que sea el camino. Habéis sido mi apoyo desde siempre. No existen palabras para agradecer todo lo que habéis hecho por mí, es imposible daros las gracias por tanta ayuda. Estaré siempre para vosotros y poder ayudaros como lo habéis hecho conmigo.

Este trabajo es tanto vuestro como mío.

¡Gracias a todos!

Índice

Capítulo 1: Introducción 24

1.1. Sensores RGB-D 27

1.1.1. Primera generación 27

1.1.2. Segunda generación 29

1.2. Aplicaciones de visión artificial de los sistemas RGB-D 31

1.3. Estructura de la Tesis Doctoral 35

Capítulo 2: Hipótesis de trabajo y Objetivos 37

2.1. Hipótesis de trabajo 39

2.2. Objetivos 39

Capítulo 3: Evaluación metrológica de dispositivos RGB-D o gaming sensors. 43

3.1. Resumen 45

Capítulo 4: Evaluación de rendimiento en aplicaciones de visión artificial .. 54

4.1. Evaluación de rendimiento y aplicación de reconstrucción 3D y reconocimiento de objetos 56

4.1.1. Resumen 56

4.2. Aplicación de visión artificial en registro de imágenes de varias modalidades 68

4.2.1. Resumen 68

Capítulo 5: Conclusiones y perspectivas futuras 81

5.1. Conclusiones 83

5.2. Perspectivas futuras 85

Referencias 88

APÉNDICE A: Indexación y factor de impacto de las publicaciones 93

APÉNDICE B: Patentes 106

APÉNDICE C: Software 165

Capítulo 1: Introducción

1. Introducción

La visión artificial es una disciplina científica cuya misión es intentar analizar y comprender las imágenes del mundo real y poder obtener información para que un ordenador sea capaz de transformar los píxeles presentes en cada imagen en información que pueda ser “comprendida” por un ordenador, tal y como los humanos usamos nuestros ojos y cerebros para comprender el mundo que nos rodea, y utilizar esa información para tareas cotidianas, como puede ser la detección de objetos. Esta comprensión se consigue gracias a distintos campos como la geometría, la física, las matemáticas y un innumerable número de campos que dependen de la aplicación de esa información adquirida. Citando a Marvin Minsky, padre de la inteligencia artificial: “*Las máquinas podrán hacer cualquier cosa que hagan las personas, porque las personas no son más que máquinas*”[1]. La visión artificial se encarga de hacer capaz a un ordenador de comprender las imágenes de la misma forma que lo hacemos los humanos. Existen numerosas aplicaciones de visión artificial. La emergencia de los sensores RGB-D (Red Green Blue – Depth) de bajo coste con una alta frecuencia de captura y con una gran resolución, ha revolucionado el mercado de los consumidores de cámaras de rango, incrementando exponencialmente el interés y capacidad de adquirir información en 3D para un gran número de aplicaciones a bajo coste. Es más, estos dispositivos han provocado un gran interés y rápida evolución de la visión artificial en 3D y en percepción de robótica, donde se han reemplazado los costosos sensores tradicionales por estos dispositivos de menor coste, que se emplazan en la familia de “gaming sensors” (este nombre lo deben de la aplicación para la que fueron desarrollados). Existen numerosas aplicaciones de visión artificial que usan esta tecnología y que están representadas en los siguientes ejemplos: reconocimiento 3D [2], reconstrucción 3D [3, 4], robótica en SLAM (Simultaneous Localization and Mapping) [5], y obviamente la industria para la que fueron creados, el entretenimiento.

1.1. Sensores RGB-D

Los sensores de profundidad incorporan a las clásicas cámaras de imágenes RGB digitales la información de profundidad para cada pixel. Conociendo los parámetros intrínsecos de la cámara RGB y la profundidad de ese pixel, se puede generar un “pixel” o punto donde se conoce la posición respecto al sensor en sus tres ejes *X*, *Y*, *Z*. Estos sensores han sufrido una revolución con los conocidos “gaming sensors”.

Actualmente existen dos generaciones de estos sensores, siendo la principal diferencia entre estas generaciones la tecnología utilizada para adquirir la profundidad. La primera generación está representada por la Kinect y Asus Xtion Pro basadas en la tecnología de luz estructurada desarrollada por PrimeSense. La segunda generación está representada por la Kinect II y está basada en la tecnología ToF (Time of Flight) y CMOS (Complementary Metal-Oxide-Semiconductor). A continuación, se explica cada una de las tecnologías utilizadas en cada una de las generaciones con más detalle.

1.1.1. Primera generación

Esta primera generación combina el principio general de la luz estructurada junto con una técnica de fotogrametría y visión computacional: la estereoscopía (depth from stereo). Con la combinación de estas dos tecnologías somos capaces de generar imágenes de rango, de forma rápida, precisa y con propiedades métricas o escala, proporcionada por la línea base del dispositivo y que es calibrado por el fabricante. Los conceptos teóricos son presentados en [6], de los que haremos un resumen a continuación.

Figura 1. Primera generación de los sensores de rango.

En relación con el principio de luz estructurada, el proyector IR (Infra-Red) proyecta un patrón pseudoaleatorio de luz infrarroja que será focalizado por la cámara IR. A través del proceso de triangulación, descrito en la *Figura 2*, y tomando como base el conocimiento del paralaje estereoscópico existente en la cámara IR que focaliza el patrón, se conseguirá determinar la imagen de profundidad (depth map), el último vértice del triángulo se consigue con la detección del patrón pseudoaleatorio en la imagen, obteniendo el punto desconocido *k*.

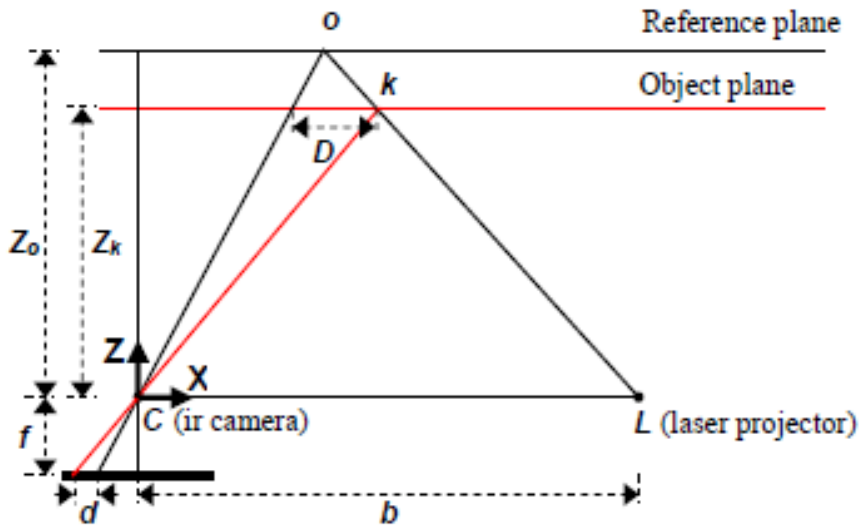


Figura 2. Principio geométrico de luz estructurada apoyado por el principio de triangulación estereoscópico.

El plano de referencia es la distancia (Z_0) en la que el patrón está en la posición perfecta (calibrada), si este patrón, se distorsiona, significa que existe algún objeto que cambia la posición de proyección de este plano. Esta distorsión en el patrón será focalizada por el plano imagen de la cámara de infrarrojos, quedando, la distorsión producida, desplazada hacia la derecha o la izquierda. A partir de ahí, y con los datos conocidos:

• *d*: desviación producida en la cámara de infrarrojos (paralaje).

• *b*: baselínea entre proyector IR y la cámara IR (conocida y calibrada).

• *f*: focal de la cámara IR (conocida y calibrada).

• z_0: distancia al plano de referencia (conocida y calibrada).

La ecuación 1 determina el valor de profundidad para el punto *k*, como sigue:

$$Z_k = \frac{Z_0}{1+\frac{Z_0}{fb}d} \quad (1)$$

despejando Z_k obtenemos el valor de profundidad proporcionado por el sensor de rango. La obtención del resto de las coordenadas X_k e Y_k , conocido como el registro entre profundidad y color, es igual para las dos generaciones de sensores de rango, con lo que se explica de forma conjunta tanto para la cámara IR (primera generación) como para la cámara RGB (segunda generación).

Las coordenadas X_k e Y_k (en las dos generaciones) están determinados por la condición de colinealidad inversa simplificada de la cámara IR o RGB (Ecuación 2):

$$X_k = (x_0 - c_x)\frac{Z_k}{f_x}$$

$$Y_k = (y_0 - c_y)\frac{Z_k}{f_y} \quad (2)$$

Donde x_0, y_0 son las coordenadas del punto en la imagen RGB o IR, c_x,c_y las coordenadas del punto principal o centro geométrico de la imagen RGB o IR y f_x, f_y la focal de la imagen RGB o IR a lo largo de los ejes *X* e *Y*, respectivamente..

1.1.2. Segunda generación

La segunda generación de estos dispositivos está representada por la Kinect 2 y utiliza la tecnología ToF (Time of Flight) o tiempo de vuelo que explicamos a continuación.

Figura 3. Segunda generación de los dispositivos de rango.

Un escáner 3D de tiempo de vuelo determina la distancia a la escena cronometrando el tiempo del viaje de ida y vuelta de un pulso de luz. Un diodo láser emite un pulso de luz y se cronometra el tiempo que pasa hasta que la luz reflejada es focalizada por un detector. Como la velocidad de la luz “*c*” es conocida, el tiempo del viaje de ida y vuelta determina la distancia recorrida por el haz, siendo dos veces la distancia entre el escáner y la superficie. Una vez que sabemos cómo funciona, explicaremos como consigue medir ese tiempo para estimar la distancia. La *Figura 4* explica de forma esquemática cómo está fabricado internamente el dispositivo y cómo funciona, el flujo de trabajo es el siguiente: se emite una señal modulada, con un desfase entre los diferentes pulsos que se emiten y se capta con la cámara para ver el desfase entre la señal modulada emitida y la recibida.

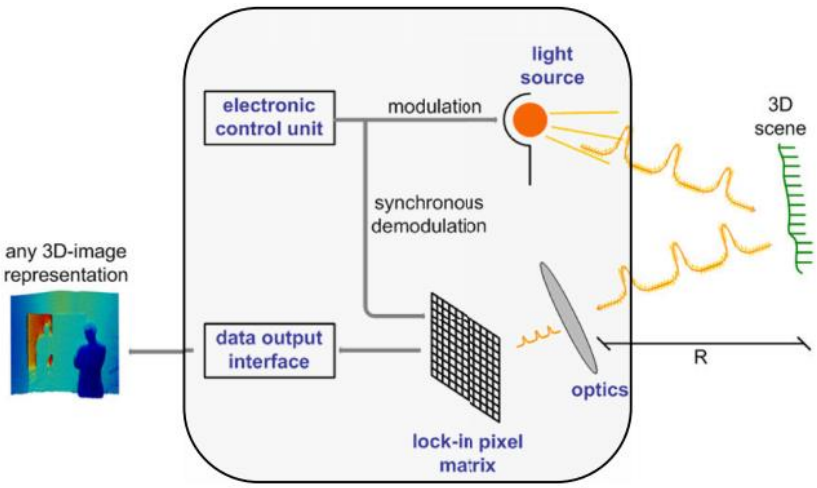


Figura 4. Diagrama interno funcionamiento de TOF CMOS extraída en web Canesta.

La modulación y cálculo de la diferencia de fase se puede encontrar de forma detallada en [7].

En los dos casos el cálculo de la profundidad es generado por un sensor o cámara distinto del sensor encargado del color. Será necesario por lo tanto un registro entre profundidad y color.

1.2. Aplicaciones de visión artificial de los sistemas RGB-D

La visión artificial o visión por computador es una disciplina científica que incluye métodos para adquirir, procesar, analizar y comprender las imágenes del mundo real y convertirlas en información que un ordenador sea capaz de interpretar.

Las principales aplicaciones son:

- Reconocimiento: Esta aplicación consiste en analizar la información de la imagen para detectar si existe o no, ciertos objetos, características o actividad desde el punto de vista semántico.
- Reconstrucción en 3D: Busca conseguir la reconstrucción del entorno o realidad en tres dimensiones a partir de diferentes imágenes.
- Análisis del movimiento: Consiste en analizar la información proporcionada por la cámara, y conseguir obtener el movimiento de la cámara o el movimiento de los objetos dentro de la imagen.
- Restauración de imagen: Esta aplicación de forma clásica estaba orientada a la eliminación de las aberraciones y problemas producidos por la cámara en la obtención de la imagen. Hoy en día esta categoría se ha visto ampliada al gran número de filtros que se utilizan en las más novedosas aplicaciones para crear una imagen “corregida” o no real.

La proliferación científica respecto a los sensores de profundidad de la primera generación ha sido muy notable, ha ayudado sobre todo al desarrollo de la visión artificial en 3D. Cabe destacar trabajos como el reconocimiento de objetos en 3D [2], el SLAM [5] o la Reconstrucción 3D [3]. Además de un gran número de trabajos que han publicado diferentes análisis[8, 9], comparaciones [10] para la primera generación de estos dispositivos, proporcionando en muchos casos “datasets” (conjuntos de datos abiertos al público para su uso) [5] para repetir los experimentos o utilizarlos en otros. El número de trabajos relacionados con la

segunda generación de estos dispositivos actualmente es muy pobre y habrá que esperar unos años para encontrar más literatura.

1.2.1 Reconocimiento de objetos

El reconocimiento de objetos es la tarea para encontrar e identificar objetos en una imagen o secuencia de video. Los humanos reconocemos una multitud de objetos en imágenes con poco esfuerzo, a pesar del hecho que la imagen del objeto puede variar un poco en diferentes puntos de vista, en diferentes tamaños o escalas, cuando están trasladados, rotados o incluso parcialmente ocultos. Esta es una tarea muy compleja y útil, en la que los dispositivos de rango han aportado una característica más, la profundidad, que puede ser de utilidad a la hora de reconocer objetos. Actualmente, el reconocimiento de objetos puede clasificarse en dos tipos de acercamientos:

- Reconocimiento con marcadores. En este tipo de reconocimiento se necesita proporcionar principalmente marcadores artificiales fácilmente detectables y que implican la colocación de los mismos en la escena u objeto.
- Reconocimiento sin marcadores. En este tipo de reconocimiento busca encontrar objetos conocidos, pero sin necesidad de marcarlos previamente, aumentando la dificultad debido al entorno menos controlado en el que se produce.

Para evitar crear un gran número de marcadores o limitarse a objetos con mucha textura, en los últimos años se han realizado grandes avances en el reconocimiento de objetos sin textura en 3D. Un gran trabajo recopilatorio de estas tecnologías ha sido realizado en [2]. A modo de resumen, existen dos grandes tipos de aproximaciones para resolver este reconocimiento. Desde una perspectiva global o local.

Perspectiva Global

En esta aproximación, se definen y reconocen objetos enteros. Lo que se busca es emparejar objetos similares entre una base de datos de objetos y un objeto en la escena de trabajo. El flujo de trabajo funciona de la siguiente forma. Se genera una base de datos que recopilará todas las hipótesis. Estas hipótesis son

generadas a partir del modelo CAD en 3D. Dichas hipótesis se describen con un algoritmo en función de sus características geométricas, que convierte ese objeto en una descripción del mismo, siendo el tamaño de esta descripción mucho menor que el objeto en sí, reduciendo así el tamaño de la base de datos. Una vez que tenemos las hipótesis empezamos a comparar estas con los datos reales. Se necesita preparar la escena para poder comparar las hipótesis con los posibles objetos existentes en la escena real, esta preparación consiste en segmentar los objetos presentes en la escena, obteniendo posibles candidatos a comparar con el objeto de referencia que se busca. De esta comparación entre objetos en la escena y las referencias se obtiene un número posible de candidatos, que deben ser aceptados o descartados. Para hacer este descarte se utiliza un algoritmo de posprocesado, como por ejemplo en el siguiente algoritmo [11].

Perspectiva Local

Esta aproximación, como su nombre indica, consiste en plantear el problema desde el punto de vista local. Se analiza la nube de puntos punto por punto, en lugar de objetos enteros. Para aumentar la eficiencia de esta aproximación se suelen escoger puntos característicos que son más fáciles de identificar. Una vez seleccionados estos puntos se necesita hacer una descripción de los mismos, lo más detallada posible y con él menor número de recursos posibles. Esto se realiza con unos algoritmos llamados descriptores, que buscan describir de forma más fidedigna posible y sin que se vea afectado por condiciones externas como rotación o escala el punto característico utilizando el menor número de recursos posibles. Una vez que se tienen los puntos definidos se comparan sus descriptores entre el modelo y la escena para encontrar puntos característicos del modelo en la escena. Esta aproximación nos evita el tramo de entrenamiento que existe en la perspectiva global. Sin embargo, al comparar un mayor número de puntos, este se vuelve más lento. Una vez realizados los emparejamientos, se buscan patrones entre el modelo y la escena que ayude a eliminar falsos positivos. Finalmente, como en el caso de los descriptores globales, se aplica un postprocesado para comprobar los emparejamientos surgidos entre las hipótesis y los candidatos.

1.2.2. Registro de imágenes

Determinar la similitud entre la información visual es necesaria en muchas tareas en la visión artificial. Más concretamente, dentro del registro de imágenes la opción más utilizada y con mejores resultados es la que basa ese registro en la búsqueda y emparejamiento de puntos que representan una característica singular en una imagen. Primero se seleccionan los puntos singulares a través de diferentes algoritmos, llamados comúnmente detectores. Una vez detectados, necesitan ser caracterizados o descritos (descriptores). El paso final de este proceso consiste en emparejar los puntos entre imágenes que presentan una mayor similitud, matching.

Sin embargo, la mayoría de las publicaciones sobre el registro entre imágenes utilizando características se reduce a las de la misma modalidad (espectro electromagnético). Existen trabajos que presentan una única solución para resolver este problema [12-14]. En [15] se realiza una evaluación de diferentes algoritmos que presentan un buen rendimiento en la descripción de características en imágenes con información de distinta modalidad. Y existe un gran número de trabajos [16, 17] que abordan el registro visto desde otra perspectiva distinto al basado en características.

1.2.3. Realidad aumentada

La realidad aumentada (RA) es una tecnológica que permite añadir a la realidad una parte sintética (generada por ordenador). La realidad aumentada es el término que se usa para definir una visión a través de un dispositivo tecnológico, de un entorno físico del mundo real, cuyos elementos se combinan con elementos virtuales para la creación de una realidad mixta en tiempo real. Una de las definiciones más aceptadas es la proporcionada por [18], en ella presenta las tres características mínimas que dotarían a un sistema de realidad aumentada:

- Combina real y virtual
- Interactivo en tiempo real
- Registrado en 3D

La visión de esta combinación de realidad e información aumentada puede ser de forma directa o indirecta, dependerá de la forma en la que el ser humano sea capaz de adquirirla.

La realidad aumentada a través de visión directa se realiza con un dispositivo que añade directamente la información virtual a nuestra visión con los dispositivos HMD (Head Mounted Display). Existen en la actualidad un gran número de prototipos creados por las grandes compañías tecnológicas que están realizando una gran inversión para llevar la realidad aumentada con visión directa a todos los consumidores. Diferentes ejemplos son: Meta [19] , Hololens de Microsoft [20], Google Glasses [21].

La realidad aumentada a través de visión indirecta se puede visualizar en cualquier pantalla y consiste en añadir información a la información captada por una cámara. Este tipo de realidad aumentada necesita de unos dispositivos más simples, con una pantalla, una cámara y un ordenador es capaz de producirse.

1.3. Estructura de la Tesis Doctoral

Esta Tesis Doctoral es presentada de acuerdo a la regulación vigente para programas de doctorado de la Universidad de Salamanca, siendo objeto de transferencia científica, a través de tres artículos publicados en revistas científicas internacionales de alto impacto, y tecnológica, representada por 2 registros de propiedad intelectual del software desarrollado y 3 patentes desarrollados durante la consecución de esta tesis. Su estructura consiste en un total de 5 capítulos acordes al desarrollo de las labores de investigación llevadas a cabo para la materialización de los objetivos fijados en la Tesis Doctoral. Se han incluido un total de tres apéndices al final del documento, con el fin de complementar el documento con información y documentación de interés.

Capítulo 2: Hipótesis de trabajo y Objetivos

2. Hipótesis de trabajo y Objetivos

En el apartado anterior se ha mostrado una visión general de los sensores de rango utilizados en esta Tesis Doctoral junto a una introducción de diferentes aplicaciones de estos dispositivos en el mundo de la visión artificial. El contexto de aplicación de las metodologías y herramientas desarrolladas en la presente Tesis Doctoral es el estudio de la utilización de estos dispositivos, así como la comprobación de la evolución del sensor en su segunda generación, tanto de forma metrológica como en aplicaciones de visión artificial. En la etapa inicial, donde se ha evaluado la viabilidad de la línea de investigación, se han recogido una serie de hipótesis de trabajo y objetivos, que marcarían la hoja de ruta a seguir de esta Tesis Doctoral.

2.1. Hipótesis de trabajo

La base de la línea de investigación es el estudio de los novedosos dispositivos de rango de bajo coste en aplicaciones de visión artificial. Estos dispositivos son capaces de capturar el entorno en 3D apoyados en diferentes tecnologías, luz estructurada y tiempo de vuelo. Estos dispositivos proporcionan una nube de puntos 3D o imagen de rango coloreada que aporta un punto extra a las cámaras tradicionales donde solo existía la radiometría de la imagen, sin tener en cuenta su profundidad. Teniendo en cuenta este nuevo punto y la existencia de dos generaciones de este dispositivo que utilizan diferentes tecnologías, se necesita conocer que tecnología y generación de dispositivos genera mejores resultados, primero desde un punto de vista metrológico, donde solo nos interesan las medidas puras, para finalmente realizar una comparación más a fondo donde se estudie su rendimiento en diferentes aplicaciones de visión artificial.

2.2. Objetivos

Enmarcados en el contexto presentado en los apartados anteriores, se plantean los objetivos de la línea de investigación materializada en esta Tesis Doctoral, clasificados en un objetivo principal y varios objetivos secundarios.

Objetivo principal:

- Analizar y evaluar el rendimiento de los sensores de profundidad para su utilización en aplicaciones de visión artificial.

Objetivos secundarios:

- Analizar y evaluar la primera y segunda generación de sensores de profundidad RGB-D desde un punto de vista metrológico.
- Analizar y evaluar el rendimiento de la primera y segunda generación de sensores de profundidad RGB-D en situaciones reales.
- Analizar y evaluar la capacidad de reconocimiento 3D con sensores de profundidad RGB-D en objetos con poca textura.
- Analizar y evaluar la precisión del posicionamiento3D (orientación y translación de un objeto concreto respeto a un sistema de coordenadas) obtenida con los algoritmos de reconocimiento 3D utilizados.
- Evaluar y analizar la capacidad de los sensores de profundidad para la reconstrucción de modelos 3D.
- Evaluar y analizar los detectores y descriptores más utilizados en el registro de imágenes de diferentes modalidades (espectros electromagnéticos), incluyendo todas sus combinaciones.
- Obtener un registro de las imágenes de diferentes modalidades haciendo un emparejamiento con los diferentes algoritmos evaluados.
- Buscar la mejora de ese registro en dispositivos mediante la realización de nuevos algoritmos de registro.

De esta forma, la línea de investigación se presenta completamente alineada con un objetivo de un proyecto de investigación desarrollado por el grupo de investigación reconocido Tidop [1, 22] y por el autor de esta Tesis Doctoral: Proyecto SICEMAM [23].

El proyecto SICEMAM consiste en el desarrollo de un sistema cibernético que permita asistir de forma eficiente a los operarios en las tareas de mantenimiento de aeronaves militares. Estas operaciones incluirán tareas de inspección, montaje y desmontaje, así como procedimientos de reparación. El gran reto de

este sistema es su vertiente más didáctica; la formación y el entrenamiento, puesto que su principal misión será la captura, gestión y transferencia del conocimiento del experto. Para ello se tendrán en cuenta las principales características de los usuarios. Asimismo, se pretende que el sistema cibernético sea escalable, sentando las bases de futuros desarrollos de I+D+i.

Capítulo 3: **Evaluación metrológica de dispositivos RGB-D o gaming sensors.**

3. Evaluación metrológica de dispositivos RGB-D o gaming sensors.

Este capítulo contiene el artículo "*Metrological comparison between Kinect I and Kinect II sensors*" publicado en la revista de alto impacto "*Measurement*" en Marzo de 2015.

3.1. Resumen

En este primer artículo se muestra una comparación metrológica entre la primera y segunda generación de los sensores Kinect I y Kinect II producidos por Microsoft. Microsoft creó estos dispositivos para integrar el entorno real dentro de los videojuegos. Para realizar esta comparación metrológica se utiliza un artefacto especial, con una forma concreta y con medidas conocidas y calibradas metrológicamente. Este consta de cinco esferas y siete cubos de diferentes tamaños. Se conoce con precisión metrológica el tamaño y posición de todos los componentes presentes en el artefacto. Estas medidas metrológicas se utilizan como verdad terreno para la comparación de los datos obtenidos por los sensores. La comparación de los dispositivos se basa en calcular la precisión y exactitud de cada uno de los dispositivos para diferentes puntos de vista respecto del artefacto metrológico. Siendo estos puntos de vista meticulosamente seleccionados con diferentes ángulos y distancias para proveer a la comparación de casos suficientemente representativos para que esta sea correcta. Los resultados de ambos dispositivos presentan una precisión similar para ambos dispositivos, siendo estos de entre 2 y 6 mm. Sin embargo, a 2 metros de distancia el sensor de primera generación presenta una desviación aproximada de 12 mm, mientras que el sensor de segunda generación permanece por debajo de los 8 mm.

En el caso de la exactitud, se mantiene la misma tendencia para la segunda generación en distancias de 1 a 2 m, siendo esta siempre menor a 5 mm negativos. En el caso de la primera generación alcanza una desviación en la exactitud del entorno de 12 mm negativos a 1 m y de 25 mm negativos a 2 m.

Con los datos del estudio de precisión realizado podemos apreciar que la precisión de los sensores de primera generación disminuye proporcionalmente

a la distancia siguiendo una función que se aproxima a un polinomio de segundo grado. Sin embargo, los sensores de segunda generación presentan unos datos más estables, pudiéndose linealizar el error cometido. El estudio de errores se ha reducido a una distancia de 4 m, debido a que es el alcance máximo de la Kinect II, y el alcance máximo recomendado por el fabricante de la primera generación, aunque este puede llegar hasta 6 m sin cometer demasiado error.

Measurement 70 (2015) 21–26

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Metrological comparison between Kinect I and Kinect II sensors

H. Gonzalez-Jorge [a,*], P. Rodríguez-Gonzálvez [b], J. Martínez-Sánchez [a], D. González-Aguilera [b], P. Arias [a], M. Gesto [b], L. Díaz-Vilariño [a]

[a] Applied Geotechnology Group, School of Mining Engineering, University of Vigo, Rúa Maxwell s/n, Campus Lagoas-Marcosende, 36310 Vigo, Spain
[b] TIDOP Research Group, Cartographic and Land Engineering Department, High Polytechnic School of Avila, University of Salamanca, 05003 Avila, Spain

ARTICLE INFO

Article history:
Received 6 September 2014
Received in revised form 7 March 2015
Accepted 20 March 2015
Available online 31 March 2015

Keywords:
Standard artefact
Optical metrology
3D sensing device
Gaming sensor
Precision
Accuracy

ABSTRACT

This work shows a metrological comparison between Kinect I and Kinect II laser scanners. The comparison is made using a standard artefact based on 5 spheres and 7 cubes. Accuracy and precision tests are done for different ranges and changing the inclination angle between each sensor and the artefact. Results at 1 m range show similar precision in both cases with values between 2 mm and 6 mm. However, at 2 m range values of Kinect I increase up to 12 mm in some cases, while Kinect II keeps all results below 8 mm. Accuracy is also better for Kinect II at 1 m and 2 m range, with values always lower than −5 mm. Accuracy for Kinect I reaches −12 mm at 1 m range and −25 mm at 2 m range. Precision study shows a decrease of precision with range according a second order polynomial equation for Kinect I, while Kinect II shows a much more stable data. Measurement range of Kinect II is limited to 4 m, while Kinect I can obtain data up to 6 m.

1. Introduction

3D modelling of the environment is something that is becoming more widespread with applicability in fields such as civil engineering [1], quality control in industry [2], robotics [3], cultural heritage [4], mining [5], or the entertainment industry [6]. In recent years, laser scanners have become widely used systems for the performance of 3D models of the environment. Depending on the type of application parameters such as range, accuracy or measurement rate are fundamental for choosing one laser scanning system over the other.

Applications in civil engineering, mining, and environmental science (e.g. surveying of a riverbank, a quarry, or a road slope) require long range (hundreds of meters) with accuracies typically around 1 cm. Architecture and cultural heritage (facades of historical buildings) require intermediate range of some tens of meters with accuracies better than 5 mm. Quality control in automotive or aerospace industry requires short range (sometimes lower than 1 m), high accuracies (0.1 mm or even better) and high measurement rate. Most of the systems used for quality control are embedded in production lines, therefore there is a need for synchronization with the manufacturing of the parts. Autonomous robots use laser scanners to map the environment, obstacle detection, and navigation-aid. They typically need medium range (30 m maximum) and low accuracy (between 3 cm and 5 cm) systems. However, since being part of a real-time control system, they need high scanning rate. Entertainment industry has more recently contributed to the development of such systems well-known as gaming sensors. They seek low-cost systems with low–intermediate ranges (between 1 m and 5 m; to work in a domestic room) and high measurement rate (to map quickly the player's movements and transmit them to the videogame). In addition, although the accuracy begins being low, the greater demands of the players, who

* Corresponding author. Tel.: +34 986818752.
E-mail address: higiniog@uvigo.es (H. Gonzalez-Jorge).

http://dx.doi.org/10.1016/j.measurement.2015.03.042
0263-2241/

Fig. 1. Kinect I laser scanner with highlighting of IR illuminator, RGB camera and IR sensor (top), and Kinect II laser scanner with RGB camera, IR sensor and IR illuminator (bottom).

want the response of the avatars accurately synchronized with their movements, is pushing the improvement of accuracy of the laser scanners.

Asus and Microsoft were two of the most popular laser scanning systems for the entertainment industry with the Xtion and Kinect systems. Kinect sold more than 24 million systems whole over the world. Both systems consist of low-cost triangulation laser scanners that have become very popular during the last couple of years. Due to the great community of potential developers working with these systems, many new applications have been developed that extend the potential of the systems to other fields different to entertainment. Some examples are indoor robotics, face recognition, virtual learning, and forensic science [7–12].

Recently, Microsoft has released Kinect II. It is based on a time-of-flight technology instead of triangulation-based former scanner. According the technical specifications, Kinect II improves Kinect I with higher camera resolution, depth resolution and frame rate [13]. However, there are not official data about the metrological characteristics of the depth measurements (i.e. accuracy and precision). These data could be very valuable for users to determine the real possibilities of the systems in many applications.

The aim of this work is to use a previously calibrated standard artefact to perform a metrological comparison between Kinect I and Kinect II sensors. Section 2 of the manuscript depicts the materials and methods used for the comparison and Section 3 the results and discussion. Conclusions are exhibited in Section 4.

2. Materials and methods

2.1. Laser scanners Kinect I and Kinect II

Main differences between Kinect I and Kinect II sensors (Fig. 1) are described in Table 1 [13]. The ranging technology of the Kinect II sensor uses a novel image system that indirectly measures the time it takes for laser pulses to travel from the IR illuminator to the image sensor after returning from the target surface. This technology divides a pixel in a half and then they are turned on or off alternatively (180° out of phase between them). The light source is pulsed in phase with the first pixel of each couple. The returned light is absorbed by the half pixel turned on and rejected by the half pixel turned off. That means that when the distance between the system and the target is increased the total amount of light absorbed by the first pixel will decrease slightly, while the second pixel increase slightly. When the target is out of range, the light photons arrive later than second halves pixels are turned on. The photons are detected by first pixels, although in another cycle.

Table 1
Technical specifications of Kinect I and Kinect II laser scanners.

	Kinect I	Kinect II
Field of view ($H \times V$)	57.5° × 43.5°	70° × 60°
Camera resolution ($H \times V$)	640 × 480 @ 30 fps	1920 × 1080 @ 30 fps (15 fps with low luminance)
Depth resolution ($H \times V$)	320 × 240	512 × 424
Maximum depth range	6 m	4.5 m
Minimum depth range	40 cm	50 cm
Depth technology	Triangulation between near infrared camera and near infrared laser source (structured-light)	Indirect time of flight
Tilt motor	Yes	No
USB standard	2.0	3.0
Supported OS	Win 7, Win 8	Win 8

Kinect II takes a first measurement with low resolution estimation, high pixel exposure time, and no ambiguities in distance. The second measurement is then taken with high precision, using the first estimate to eliminate any ambiguities.

Kinect II has built in ambient light rejection, where each pixel individually detects when that pixel is over saturated with incoming ambient light and it resets the pixel in the middle of an exposure. On the contrary, Kinect I does not provide ambient light rejection.

2.2. Metrological comparison

Metrological comparison between Kinect I and Kinect II was done by using a standard artefact developed at University of Vigo [14–16]. This artefact (Fig. 2) consists of five delrin spheres of nominal diameter 100 mm equidistantly assembled on an aluminium block and seven cubes of edge dimensions 100 mm, 80 mm, 60 mm, 40 mm, 30 mm, 20 mm, and 10 mm. The artefact was calibrated in an ENAC accredited laboratory according ISO 17025:2005 using a coordinate measurement machine. Calibration is yearly updated and no significant changes have appreciated to work with laser scanning systems with precisions above mm. Metrological characteristics of the artefact are shown in the bibliography, in the same way that the results obtained by the sensor Kinect I. These results will be used in the present work to compare with the data obtained with Kinect II.

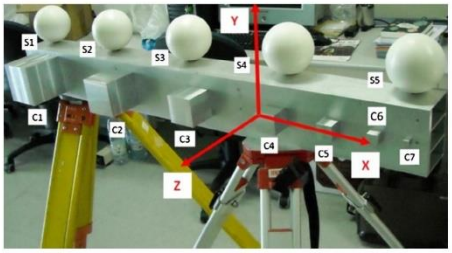


Fig. 2. Standard artefact used for metrological comparison between Kinect I and Kinect II.

The metrological evaluation performed with Kinect II sensor follows the procedure used with the same standard artefact and Kinect I [6]. It basically consists of placing the artefact and the Kinect sensor on two surveying tripods and performing six complete measurements at 1 m and 2 m range, with 45°, 90° (coincident with Z axis in Fig. 1), and 135° angles between artefact and sensor (Fig. 3). Depth measurements were also made at 90° angle, from 1 to 4 m. For ranges higher than 4 m Kinect II appears out of range. Kinect II data are stored using the Microsoft Kinect SDK for range measurements. The whole data acquisition was done indoors where the illumination consisted of fluorescence tubes. Example of the point cloud is show in Fig. 4.

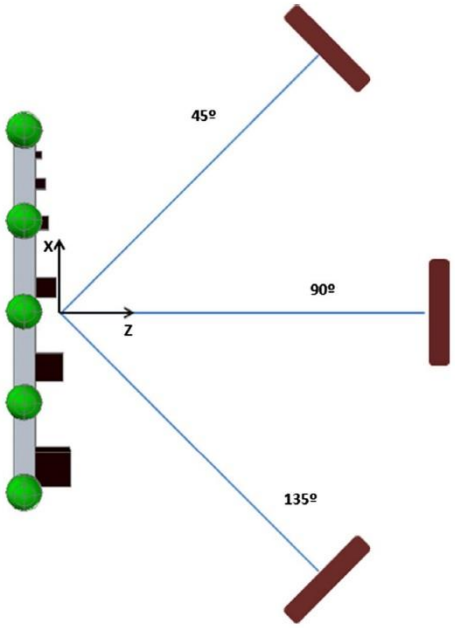


Fig. 3. Scheme of data acquisition.

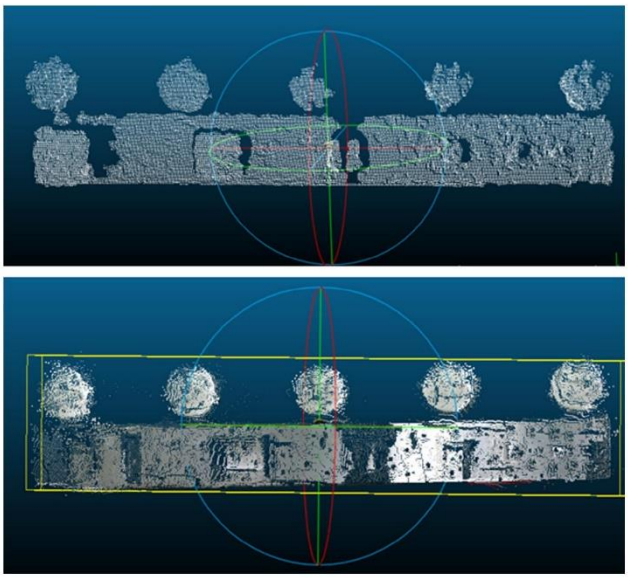


Fig. 4. Point cloud from standard artefact obtained with Kinect I laser scanner (top) and Kinect II laser scanner (bottom).

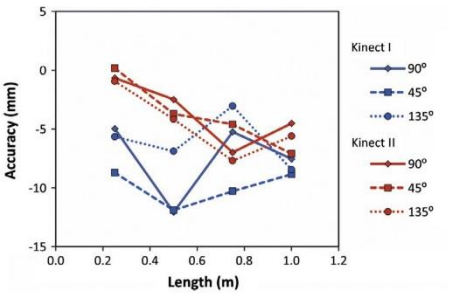


Fig. 5. Accuracy for 1 m range.

Accuracy *acc* is calculated as the difference between the distance values obtained for the standard artifact D_i^{SA} and the Kinect I $D_i^{\text{Kinect I}}$ (Eq. (1)) or II $D_i^{\text{Kinect II}}$ (Eq. (2)). Eqs. (3)–(5) show as the distance values are obtained from the center of each sphere $(x_{i+1}, y_{i+1}, z_{i+1})$ in relation with sphere 1 (x_1, y_1, z_1); i values range from 1 to 4. The coordinates from the center of the spheres are obtained in all cases using a least square fitting algorithm, taking into account the data from the coordinate measurement machine (standard artifact data) or the data from the point clouds (Kinect I or II).

$$acc_i^{\text{Kinect I}} = D_i^{SA} - D_i^{\text{Kinect I}} \tag{1}$$

$$acc_i^{\text{Kinect II}} = D_i^{SA} - D_i^{\text{Kinect II}} \tag{2}$$

$$D_i^{SA} = \sqrt{(x_{i+1}^{SA} - x_1^{SA})^2 + (y_{i+1}^{SA} - y_1^{SA})^2 + (z_{i+1}^{SA} - z_1^{SA})^2} \tag{3}$$

$$D_i^{\text{Kinect I}} = \sqrt{(x_{i+1}^{\text{Kinect I}} - x_1^{\text{Kinect I}})^2 + (y_{i+1}^{\text{Kinect I}} - y_1^{\text{Kinect I}})^2 + (z_{i+1}^{\text{Kinect I}} - z_1^{\text{Kinect I}})^2} \tag{4}$$

$$D_i^{\text{Kinect II}} = \sqrt{(x_{i+1}^{\text{Kinect II}} - x_1^{\text{Kinect II}})^2 + (y_{i+1}^{\text{Kinect II}} - y_1^{\text{Kinect II}})^2 + (z_{i+1}^{\text{Kinect II}} - z_1^{\text{Kinect II}})^2} \tag{5}$$

Precision of Kinect II is evaluated using the standard deviation of least square algorithm applied to the fitting of the spheres and the top face of the largest cube of the artifact (100 mm edge) [6,7,14,17–19]. In particular, the procedure for estimating precision included: the same measurement procedure, the same observer, the same measuring instrument, used under the same conditions, the same location, and repeated measurements from the same measurand. To this end, Eq. (6) shows a quantitative parameterization of the results based on the standard deviation of fitting residuals d_i, where N depicts the number of points from the cloud that are used for the fitting. Low precision will be indicative of a noisy point cloud that produces large d_i values.

$$prec = \sqrt{\frac{\sum_{i=1}^{N} d_i^2}{N-1}} \tag{6}$$

3. Results and discussion

Figs. 5 and 6 show the accuracy results for Kinect I and Kinect II sensors for ranges of 1 m and 2 m, with angles

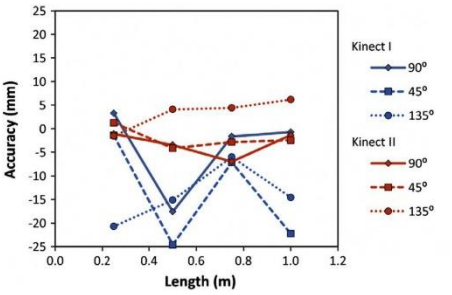


Fig. 6. Accuracy for 2 m range.

between the laser scanners and the standard artifact of 45°, 90°, and 135°. Kinect I depicts accuracy values ranging between −2 mm and −12 mm for 1 m range and between 4 mm and −25 mm for 2 m range. A clear decreasing of accuracy with range is shown. Kinect II shows accuracy values between 0.1 mm and −7.5 mm for 1 m range and between 5 mm and −7 mm for 2 m range. The accuracy decreasing with range is much less pronounced. Accuracy values of Kinect II clearly improve the values obtained for Kinect I at 2 m range. In both cases there is not a trend between the accuracy and the angle between the sensor and the standard artifact.

Fig. 7 depicts the precision results for 1 m range and Fig. 8 the results for 2 m range. Both sensors show a decreasing of precision with the increasing of range, although it is much more prominent in sensor Kinect I. Results for 1 m range are very similar for both sensors. The precision values range between 1.5 mm and 6 mm. However, for 2 m range, Kinect II clearly improves the Kinect I data. Kinect II shows precisions values lower than 8 mm, while Kinect I shows values over 10 mm in many cases.

The incidence angle does not affect the obtained results, as well as it does not influence precision or accuracy values.

The accuracy/precision ratio is for all cases approximately between 1 and 2. This result correlates with that

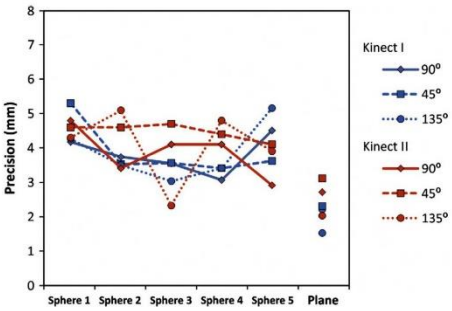


Fig. 7. Precision for 1 m range.

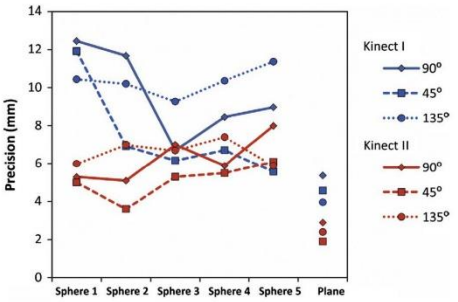


Fig. 8. Precision for 2 m range.

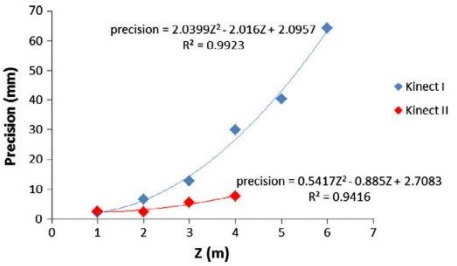


Fig. 9. Precision trend with range.

obtained for commercial systems (i.e. Faro Focus 3D [20]) with precision values between 0.6 and 2.2 mm and accuracy of 2 mm. In this case the ratio depicts values between 0.9 and 3.3.

Fig. 9 exhibits the precision as a function of range for the laser scanners Kinect I and II. Precision was obtained from the standard deviation of the least squares plane fitting to the top face of the largest cube (90° angle). Precision data for 1 m and 2 m range come from those depicted for the planes in Figs. 7 and 8. Precision data for 3 m and larger ranges are specifically calculated to this part.

Kinect I shows a measurement range between 1 m and 6 m, while Kinect II shows it between 1 m and 4 m, both in agreement with the technical specifications. Kinect I shows a precision decreasing in agreement with a second order polynomial [17], while Kinect II does not show this clear mathematical behavior. Precision for Kinect II appears much more stable with the increasing of range inside the measurement window. This fact could be quantified with the fitter second order polynomial that shows a lower Z^2 coefficient for Kinect II (0.54 vs. 2.04 in Kinect I).

4. Conclusion

This works shows the comparison between Kinect I and Kinect II laser scanners using a standard artefact. The comparison evaluates the accuracy and precision at angles of 45°, 90° and 135° and distances of 1 m and 2 m. Precision

was also evaluated for larger ranges up to 6 m for Kinect I and 4 m for Kinect II.

Results at 1 m range show similar precision values in both laser scanners; however at 2 m range values of Kinect II improve the results of Kinect I. Accuracy shows a similar pattern, although even for 1 m range values for Kinect II appear slightly better.

Precision was tested until the range limit of the sensors. Precision obtained for Kinect I sensor decreases following a second order polynomial equation, while Kinect II fits a more stable behaviour. Decreasing of precision with range for Kinect II is less appreciable. Range limit of Kinect II appears at 4 m, while Kinect I achieves 6 m.

The metrological comparison performed in this work concludes that Kinect II could be useful for the same technical applications as Kinect I (robotics, quality control of low-tolerance parts, or indoor mapping), improving accuracy and precision. It must also be noted the stability of measurements with range that could be very valuable for many uses, the same way the better performance in outdoor environments.

Acknowledgements

Authors want to give thanks to the Xunta de Galicia (Grant Nos. IPP055 – EXP44, EM2013/005, CN2012/269, R2014/032) and Spanish Government (Grant Nos. TIN2013-46801-C4-4-R, ENE2013-48015-C3-1-R, SICEMAM IPT-2012-1121-370000).

References

[1] H. González-Jorge, I. Puente, B. Riveiro, J. Martínez-Sánchez, P. Arias, Automatic segmentation of road overpasses and detection of mortar efflorescence using mobile LiDAR data, Opt. Laser Technol. 54 (2013) 353–361.

[2] T. Luhmann, F. Bethmann, B. Herd, J. Ohm, Comparison and verification of optical 2D surface measurement systems, Int. Arch. Photogramm., Rem. Sens. Spatial Inform. Sci. XXXVII –B5 (2008) 51–56.

[3] U. Weiss, P. Biber, Plant detection and mapping for agricultural robots using a 3D LiDAR sensor, Robot. Auton. Syst. 59 (5) (2011) 265–273.

[4] B. Riveiro, P. Morer, P. Arias, I. de Arteaga, Terrestrial laser scanning and limit analysis of masonry bridges, Constr. Build. Mater. 24 (4) (2011) 1726–1735.

[5] J. Armesto, C. Ordoñez, L. Alejano, P. Arias, Terrestrial laser scanning used to determine the geometry of a granite Boulder for stability analysis purposes, Geomorphology 106 (3–4) (2009) 271–277.

[6] H. González-Jorge, B. Riveiro, E. Vázquez-Fernández, J. Martínez-Sánchez, P. Arias, Metrological evaluation of Microsoft Kinect and Asus Xtion sensors, Measurement 46 (6) (2013) 1800–1806.

[7] H. González-Jorge, S. Zancajo, D. González-Aguilera, P. Arias, Application of Kinect gaming sensor in forensic science, J. Foren. Sci. 60 (1) (2015) 206–211.

[8] W. Jia, W.J. Ji, J. Saniie, E. Oruklu, 3D image reconstruction and human body tracking using stereo vision and Kinect technology, in: IEEE International conference on Electro-Information Technology, 2012, pp. 1–4. (Article number 6220732).

[9] C.C. Martín, D.C Burkert, K.R. Choi, N.B. Wieczorek, P.M. McGregor, R.A. Herrmann, P.A. Beling, A real time ergonomic monitoring system using the Microsoft Kinect, in: IEEE Systems and Information Engineering Design Symposium (2012); pp. 50–55. (Article number 6215130).

[10] G. Csaba, L. Somlyai, Z. Vamossy, Differences between Kinect and structured lighting sensor in robot navigation, in: IEEE International Symposium of Applied Machine Intelligence and Informatics (2012); pp. 85 – 90. (Article Number 6208934).

[11] T. Dutta, Evaluation of the Kinect sensor for 3D kinematic measurement in the workplace, Appl. Ergon. 43 (4) (2012) 645–649.

[12] F. Cassola, L. Morgado, F. de Carvalho, H. Paredes, B. Fonseca, P. Martins, Online-Gym: a 3D virtual gymnasium using Kinect interaction, Prcedia Technol. 13 (2014) 130–138.

[13] Kinect II Tech Specs. <http://123kinect.com/everything-kinect-2-one-place/43136/>.

[14] H. González-Jorge, B. Riveiro, J. Armesto, P. Arias, Standard artifact for the geometric verification of terrestrial laser scanning systems, Opt. Laser Technol. 43 (2011) 1249–1256.

[15] H. González-Jorge, B. Riveiro, J. Armesto, P. Arias, Verification artifact for photogrammetric measurement systems, Opt. Eng. 50 (7) (2011) 073603. 1 – 8.

[16] I. Puente, H. González-Jorge, B. Riveiro, P. Arias, Accuracy verification of the Lynx mobile mapper system, Opt. Laser Technol. 45 (2013) 578–586.

[17] K. Khoshelham, S.O. Elberink, Accuracy and resolution of Kinect depth data for indoor mapping applications, Sensors 12 (2012) 1437–1454.

[18] G. Guidi, M. Russo, G. Magrassi, M. Bordegoni, Performance evaluation of triangulation based range sensors, Sensors 10 (2010) 7192–7214.

[19] BIPM, Evaluation of measurement data – guide to the expression of uncertainty in measurement, JCGM 100:2008.

[20] Faro Focus 3D datasheet. <www2.faro.com/site/resources/share/944>.

Capítulo 4: Evaluación de rendimiento en aplicaciones de visión artificial

4. Evaluación de rendimiento en aplicaciones de visión artificial

Este capítulo contiene dos artículos que serán analizados en dos secciones.

4.1. Evaluación de rendimiento y aplicación de reconstrucción 3D y reconocimiento de objetos

La primera sección de este capítulo contiene el artículo “*Analysis and Evaluation between the First and the Second Generation of RGB-D Sensors*” publicado en la revista de alto impacto IEEE “*Sensors Journal*” en Julio de 2015.

4.1.1. Resumen

En este segundo artículo realizamos un análisis en mayor profundidad de los dos dispositivos, se realiza un análisis de rendimiento utilizando aplicaciones de visión artificial comunes. La idea de escribir este artículo ha sido propiciada por la introducción de estos dispositivos de videojuegos, y en concreto con la irrupción de la segunda generación, el bien conocido Kinect II, que ha producido que la interacción entre estos sensores y las aplicaciones de ingeniería inversa y visión artificial se haya visto reforzada. Este nuevo sensor basado en una nueva tecnología ToF (Time of Flight o tiempo de vuelo) diferente de la primera generación de los dispositivos RGB-D, que está basado en luz estructurada (Asus Xtion pro, Kinect I o Primesense Carmine). A pesar de que la segunda versión de este dispositivo tiene mejores características técnicas no podemos predecir cómo se traducirán estas mejoras en diferentes aplicaciones de visión artificial o ingeniería inversa. Este trabajo busca comparar de forma cuantitativa el rendimiento de estas dos generaciones de dispositivos en función de dos escenarios aplicados: reconstrucción 3D y detección de objetos en 3D. En este artículo se han realizado varios trabajos. Primero se introduce un novedoso “dataset”, definiendo una verdad terreno de alta precisión obtenida con un láser metrológico de alta precisión. Tener este “dataset” nos permite realizar dos análisis: (i) análisis de rendimiento en términos de precisión en la reconstrucción y (ii) una comparación en términos de reconocimiento de objetos 3D y en la estimación del posicionamiento de los mismos. Los resultados obtenidos confirman que la nueva versión del sensor Kinect ha obtenido mayor precisión y

menos ruido en la aplicación de reconstrucción 3D. Es más, se proporciona una estimación cuantitativa de cuanto mejoran los resultados en el reconocimiento 3D tomando como hipótesis una mayor precisión y exactitud en la adquisición de la nube de puntos por parte de la segunda generación de estos dispositivos.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/JSEN.2015.2459139

Analysis and Evaluation between the First and the Second Generation of RGB-D Sensors

M. Gesto-Diaz, F. Tombari, P. Rodriguez-Gonzalvez and D. Gonzalez-Aguilera

***Abstract*—With the recent introduction of the new Kinect II, the second generation of the well-known Microsoft Kinect sensors, the connection between RGB-D sensors, reverse engineering and computer vision applications is reinforced. This new sensor is based on a Time-of-Flight (ToF) technology, which differs from the previous generation of RGB-D sensors, including other devices such as the Asus Xtion Pro and PrimeSense Carmine, which was based on structured light. Although characterized by better technical specifications, this does not neccessarily translate to the improvemens in its application tasks. This work aims at comparing quantitatively the Kinect II with respect to the first generation of RGB-D sensors in terms of two specific application scenarios: 3D reconstruction and object recognition. To this end, we propose a novel dataset with ground truth obtained with a metrological laser scanner, which allows a two-fold analysis: (i) a performance comparison in terms of reconstruction accuracy and (ii) a comparison in terms of object recognition and 3D pose estimation. The obtained results confirm that the new version of the Kinect sensor demonstrate higher precision and less noise under controlled conditions. Furthermore, we provide a quantitative estimation of how much such factors turn out into an improvement in terms of object recognition rate and 3D pose estimation.**

***Index Terms*—RGB-D sensor, accuracy assessment, object recognition, gaming sensors, Kinect II and Asus Xtion Pro**

I. Introduction

THE introduction of affordable, high frame-rate and dense RGB-D sensors such as Microsoft Kinect, Asus Xtion Pro and PrimeSense Carmine has revolutionized the market of consumer depth cameras, increasing exponentially the interest and the deployment of 3D data for a higher number of applications. Indeed, these devices have notably pushed forward the research activity in 3D computer vision and robotic perception, and they have now replaced traditional sensors for applications such as 3D object recognition [1], Simultaneous Localization and Mapping (SLAM) [2] and 3D reconstruction [3]–[5]. Due to the popularity obtained by this first generation of RGB-D sensors, several research works were recently published with the aim of analyzing and comparing these sensors in terms of accuracy of the acquired 3D data. At the same time, several benchmark datasets acquired with such devices were proposed in literature, aimed at the aforementioned applications, as well as surveys and performance evaluations among different techniques tested on such benchmarks [2], [6].

Recently, the Kinect II sensor was released on the market by Microsoft, based on Time-of-Flight (ToF) technology acquired by Microsoft from Canesta in 2010. The most evident innovation of this sensor with respect to previous generation is the use of a different 3D sensing technology, currently based on ToF. According to the technical specifications, Kinect II improves the previous generation of structured light sensors with a higher color camera resolution and the possibility to work outdoors. Nevertheless, it is not possible to directly translate and quantify these more advanced characteristics into benefits in terms of the final application, especially due to the switch in the sensing technology from structured light to ToF.

For this reason, this work proposes a performance evaluation aimed at comparing Kinect II with previous RGB-D sensors in terms of accuracy of the acquired 3D data. The goal is to characterize the quality of the acquired 3D data and to quantify the differences between these sensors, by means of a novel dataset, where each scene was acquired by both types of sensors, as well as by a metrological laser scanner which provides the ground-truth for the quantitative comparison. Indeed, we wish to point out that several authors have coped with metrological assessments of the Kinect sensor, in terms of accuracy analysis [8]; calibration, resolution and set up analysis [9] and performance evaluation under different computer vision applications [10]. At this moment the number of works with Kinect II is reduced [11], but for similar devices (Mesa Imaging, PMD and Optrima) based on ToF technology there are some works related to the set-up, accuracy and practical results [12]–[14].

In addition, the present manuscript focuses on some of the most popular applications of RGB-D sensors, 3D object recognition and pose estimation, and it aims to compare the two generations of sensors also in terms of recognition rate and 3D pose estimation accuracy, still exploiting the novel dataset introduced with this work and its ground-truth. By comparing the performance yielded by state-of-the-art 3D object recognition algorithms, we aim to assess the advantages brought in by Kinect II sensor directly in terms of application, so to better analyze the impact that Kinect II could provide to the community. In this case, there is yet no research works that has coped with quantitatively benchmarking 3D object recognition and pose estimation algorithms on 3D data acquired with Kinect II.

This paper is structured as follows: after this Introduction motivating the advances of gaming sensors, section II presents the materials used to create the dataset and the

M. Gesto-Diaz is with the University of Salamanca, Ávila 05003, Spain (e-mail: mgesto@usal.es).

F. Tombari is with DISI, University of Bologna. V.le del Risorgimento 2 Bologna (Italy)(e-mail:federico.tombari@unibo.it)

P. Rodriguez-Gonzalvez is with the University of Salamanca, Ávila 05003, Spain (e-mail:pablogrsf@usal.es).

D. Gonzalez-Aguilera is with the University of Salamanca, Ávila 05003, Spain (e-mail:daguilera@usal.es).

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/JSEN.2015.2459139

methods developed to acquire and compare the dataset. The experimental results are reported and discussed in section III, while final remarks and conclusions are drawn in section IV.

II. Materials and Methods

A. Sensors

Two RGB-D sensors are being evaluated, the Asus Xtion Pro , as a representative of the first generation of RGB-D sensors based on Primesense structured light technology, and the Microsoft Kinect II sensor, based on ToF technology [15]. A comparison in terms of their technical features is reported in Table I and represented in (Figure 1).

Fig. 1. Asus Xtion Pro (Left) and Microsoft Kinect II (Right).

TABLE I
Comparison between technical specifications of both sensors.

	Asus Xtion Pro	Kinect II
Color Camera	640x480x24 bpp 4:3RGB at 30 fps	1920x1080x16 bpp 16:9 YUY2 at 30 fps
Depth Camera	640x480x16 bpp 13-bit depth	512x424x16 bpp, 13 bit depth
Max. depth distance	8 m	4.5 m
Min. depth distance	0.50 m	0.50 m
Latency	90 ms with processing	60 ms with processing
Field of view	57.5° Horizontal and 43.5° Vertical	70° Horizontal and 60° Vertical

Since the technology related to the first generation of RGB-D sensors is well-known, it has been analysed in several papers [8], [9]. As for Kinect II, being a commercial solution whose characteristics have only been partially disclosed, it is yet not possible to know the details about the technology employed. But to aid the reader to get a general idea of the device measuring principle, we present one of the most plausible theories, according to [11], which states that the Kinect II uses a phase shift with three different wavelengths.

In this paper, the accuracy assessment of both sensors is validated with a metrological laser scanner, Hexagon Metrology Absolute Arm 7325SI, which provides the ground truth of the dataset. This sensor is based on the principle of optical triangulation. The main technical specifications of this metrological arm are shown in Table II. A example of the precision achieve can be observed in the Table II with the deviations in a sphere in the field Scanning System Verification Report.

TABLE II
Technical specifications of the laser scanner Hexagon Absolute Arm 7325SI.

Hexagon Metrology Absolute Arm 7325SI	
Measuring range	2.5 m
Probing point repeatability	±0.049 mm
Probing volumetric accuracy	±0.069 mm
Scanning system accuracy	±0.042 mm
Integrated scanner RS	
Maximum point acquisition rate	50000 points/s
Points per line	1000
Line rate	50 Hz
Point spacing (min)	0.046 mm
Accuracy (2 sigma)	30 μm
Certificate of Traceable Calibration(ASME B89.4.22-2004)	
Point repeatability achieved	0.017 mm
Length accuracy achieved	0.022 mm
Scanning System Verification Report (Sphere @ 80%)	
Max	0.0233 mm
Standard deviation	0.0167 mm
Diameter deviation	-0.0310 mm

B. Dataset

Four categories of objects were designed to acquire a dataset aimed at assessing the accuracy of RGB-D sensors for 3D reconstruction and object recognition applications(Figure 2). These categories were chosen using the most common categories presented in benchmark datasets for 3D object recognition (i.e. basic primitives and objects with different shapes and sizes), including a category for textured objects. From a performance point of view, each model encloses a local reference frame placed in the centroid of the object, which defines the object's pose. This local reference frame is used to compute a dimensional analysis of the object, as well as its pose estimation (6 DOF-Degrees of Freedom) in the scene and thus to assess its object recognition capabilities. In addition, trying to reinforce the accuracy assessment for object recognition, three different types of scenes have been generated: (i) scenes with one single object; (ii) scenes with several objects belonging to the same category and (iii) scenes which mixed objects of different categories.

Overall, the dataset consists of 24 CAD models grouped in 4 categories and 50 scenes. A first subset of scenes (1-24) includes only one model. Note that, for what concerns the basic shapes, there are 3 single-model scenes because, in this case, we have 3 scale factors (0.5-0.75 and 1) for each model. Then, there are 5 scenes for each category with mixed models belonging to it. These categories are Basic shape (25 to 29), Household objects (30 to 34), Toys (35 to 39) and Textured objects (40 to 44). Finally, the dataset contains 6 scenes with models of all categories (45 to 50). To help understanding the structure of the proposed dataset, Figure 2 summarizes the models together with the corresponding scenes where they appear. In Figure 2, the model name together with the group category are reported in black, while the corresponding scene numbers are depicted in light blue.

It should be remarked that these models were generated from their original 3D CAD model by means of a 3D printer, BQ WitBox [16], with a precision of ±0.1 mm according to manufacturer specifications.

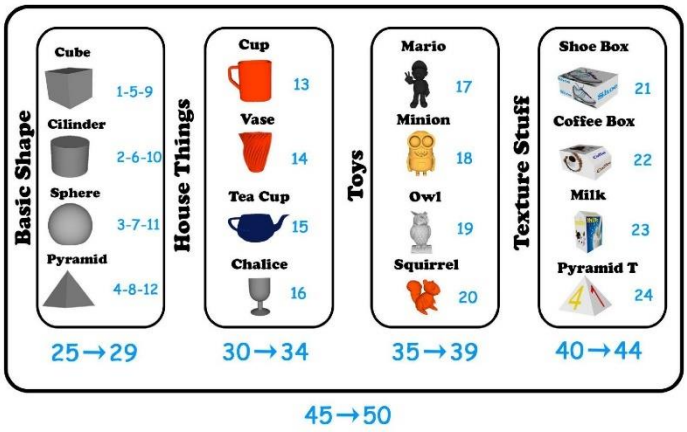


Fig. 2. Novel dataset designed for the accuracy assessment of gaming sensors in reverse engineering and object recognition applications. This dataset encloses a ground truth provided by a metrological laser scanner.

C. Data acquisition

The protocol developed for data acquisition should guarantee enough precision to allow assessing the accuracy of Asus Xtion and Kinect II from a 3D reconstruction and object recognition point of view. Therefore, the different sensors used, including the own metrological laser scanner, should stay in the same position during data acquisition. Since it is almost impossible to guarantee that the three sensors have the same position due to their own intrinsic features (i.e. different baseline between RGB and depth cameras, different field of view, different resolution, different components), we decided to put sensors as close as possible in a fixed and stable position, allowing their registration (i.e. sensor position) and the acquisition of the different scenes simultaneously (Figure 3). To solve the registration between the sensors, a special scene was designed with boxes of different sizes oriented along different directions. Furthermore, in these planes several control points have been established for both devices. Finally, a manual registration between these control points in the scene captured with Asus Xtion Pro and Kinect II and the ground truth provided by the metrological laser is performed and refined by applying the Iterative Closest Point (ICP) algorithm [17].

Once the registration between the three sensors is obtained, the dataset scenes were acquired with the metrological laser scanner (so to generate the ground truth), as well as with the two RGB-D sensors. The accuracy of the metrological laser scanner for estimating the 6-DOF (x,y,z, roll(α), pitch(β), yaw(γ)) pose of each object is estimated in the worst case as ± 0.5 mm, based on the analysis of the discrepancies of the original CAD model and the registered point clouds from metrological laser scanner.

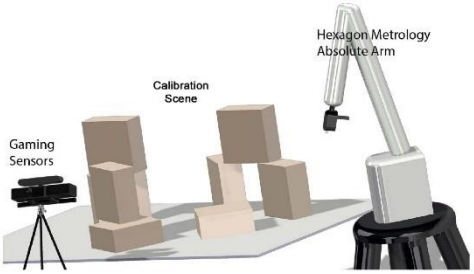


Fig. 3. Configuration of sensors to acquire the scenes.

D. Comparison in terms of 3D data accuracy

Before establishing a performance comparison between the two sensors, the error distribution of the acquired data must be computed in order to determine whether the data follows a normal or non-normal distribution. This is carried out in Subsection II-D1. Once the dataset has been captured and the normality of the dataset has been determined, a comparison in terms of 3D accuracy is performed. In particular, this accuracy assessment is carried out from both a quantitative and qualitative point of view. The former refers to the statistical accuracy in terms of non-parametric approaches (described in subsection II-D2), whereas the latter is related with the completeness of the object modeled by means of its number of fitted points and its color misalignment, and it is illustrated in subsections II-D3 and II-D4.

1) Analysing the normality of dataset : Two specific statistical procedures have been applied in assessing whether a sample of independent observations follows a normal or

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/JSEN.2015.2459139

non-normal distribution: graphical methods (QQ-plots) and numerical methods based on the Anderson-Darling normality test [18] [19]. Although the sensitivity of normality tests to non-normal data could seem an efficient alternative, it should be remarked that these tests do not work properly with large datasets since the central limit theorem comes into play [20], so that normality tests were only applied in those cases with a reduced number of observations. For large datasets, a better diagnostic to check a deviation from the normal distribution is the visual plot quantile-quantile (QQ-plot). In this case, the quantiles of the empirical distribution function are plotted against the theoretical quantiles of the normal distribution. If the distribution follows a Gaussian function, the QQ-plot should be a diagonal straight line. Furthermore, the distribution of errors can be visualized by a histogram of the errors, where the number of errors (frequency) within certain predefined intervals is plotted together with the theoretical curve for a normal Gaussian distribution. However, the histograms are not used as normality criteria, since their shape change according to the bin size. In the experimental results, both devices presents a non-normal distribution for the discrepancies between the model obtained for the device and the ground truth.

In our case, to analyse the normality of dataset the error associated to each of the two evaluated RGB-D sensors has been computed based on the orthogonal distance between the point clouds provided by gaming sensors and the surface scanned with the metrological arm. In particular, this orthogonal distance is assessed in terms of discrepancy (difference) as follows (Equation 1):

$$Dif = S_{GT} - P_{GS} \tag{1}$$

where S_{GT} represents the surface (mesh) provided by the metrological laser scanner which performs as ground truth, and P_{GS} represents the point cloud provided for each RGB-D sensor.

2) Statistical accuracy : If a normal distribution can be assumed and no outliers are presented in the dataset, the classical statistical accuracy measures based on mean error and standard deviation can be applied. On the contrary, if the Anderson-Darling normality test together with graphical QQ-plots of the errors reveal an excessive amount of outliers, another approach for deriving accuracy measures has to be employed. Such an approach has to be robust to outliers, and the probability assumptions to be made should not assume normality of the error distribution. Our proposal in this case is to apply a non-parametric bootstrap strategy [21] supported by robust estimators (median and biweight midvariance-$BWMV$) considering a random subsample of the differences computed in the step above (Equation 1).

The basic idea of bootstrap involves inferring the variability of an unknown distribution from which your data are drawn by re-sampling with replacement from a sample of independent observations (Dif_1,,Dif_n) where Dif_1 denotes the first and Dif_n the last of the differences. The main goal is to determine if the ground truth discrepancies in a sample of n points are significant with a confidence level of $1 - \alpha$. The bootstrap method proposed consist on a statistical re-sampling method that analyses the variability of an estimator E obtained from a sample Dif = Dif_1,..., Dif_n and follows an F distribution. The selection of the re-sampling method is important in order to in achieve a distribution of E^* that properly approximates the distribution of F. In our case, the bootstrap method has been adapted in the following way:

1) The orthogonal difference between the ground truth and the point clouds acquired by the RGB-D sensors are obtained. This difference gives an idea of the accuracy of each RGB-D sensor.

2) The bootstrap samples are generated randomly from the differences among the n original data points using a large number of repetitions, $b = 1, ..., B$. The corresponding values of the differences, b, are obtained. Based on its replacement criteria, it is possible that some of the original data points are presented more than once, and others are omitted.

3) For each bootstrap replicas the samples quantiles $Q(p)$ together with the robust estimators (median and $BWMV$) are estimated using a large number of repetitions, $b = 1, ..., B$, so that we obtain a collection of b estimates of the accuracy. In particular, first ($Q_{25\%}$) and third ($Q_{75\%}$) quantiles together with the interquartile range (IQR=$Q_{75\%}$-$Q_{25\%}$) are estimated. In addition, these quantiles are complemented with a robust estimator for the central tendency and the dispersion based on the median,m , and the $BWMV$ function [22], respectively (Equations 2 and 4).

$$m = \begin{cases} x_{\frac{n+1}{2}} & if\ n\ is\ odd \\ \frac{1}{2} \cdot \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & otherwise \end{cases} \tag{2}$$

$$U_i = \frac{x_i - m}{9MAD} \tag{3}$$

$$BWMV = \frac{n \sum_{i=1}^{n} a_i (x_i - m)^2 (1 - U_i^2)^4}{\left(\sum_{i=1}^{n} a_i (1 - U_i^2)(1 - 5U_i^2)\right)^2} \tag{4}$$

where m is the median, n is the number of points, U is a parameter from Equation 3, x_i is the mean value and MAD is the Median Absolute Deviation, defined as:

$$MAD = median(|Dif_i - m_{dif}|) \tag{5}$$

where Dif denotes the orthogonal distances and m_{dif} is the median of the differences in orthogonal distances. Finally, the value of the parameter a (Equation 4) could be 0 or 1 depending on the U value. If $-1 \leq U \leq 1$, thus a = 1; in any case a = 0. The essential idea of bootstrap is that this set of quantiles and robust estimators provide a distribution that approximates the real distribution of the actual dataset.

4) Finally, a bias-corrected estimate of the parameter can be computed by averaging the b different values. The central tendency and dispersion of the estimator

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/JSEN.2015.2459139

E^* can be inferred by computing the median and the $BWMV$ function of this collection of random samples, whereas the confidence interval on the accuracy can be approximated by using the appropriate upper and lower percentiles of the observed bootstrap sample quantiles and robust estimators, that is, the percentile bootstrap confidence intervals (PBCI).

3) Spatial completeness of the object : A simple comparison based on the number of points fitted for each model is established as a measure of its completeness. It is important to know how many points of each model can be obtained and which RGB-D sensor is more suitable, as well as the number points that are capable to acquire of each element. Considering that the two RGB-D sensors have different technical specifications in terms of field-of-view and number of pixels (Table I), it is expected that the Kinect I will have more fitted points. Although we have no certain knowledge of how commercial sensors work internally, we can assume that the PrimeSense based cameras use some form of correlation to compare the observed pattern to a reference pattern. Any form of correlation over an area of the image inevitably creates a dependency in-between pixels and has a low-pass or smoothing effect. On the other hand, for a ToF camera we can assume that the depth is estimated independently for each pixel using a variant of phase-shifting. Therefore, it might occur that, even if the number of pixels is lower on a ToF camera, the actual precision is better. This can be observed in a simple Kinect I vs. Kinect II test, when observing the details on the fingers of an projected hand.

To prove this, the protocol presented in [23] is followed. Where a special artifact based on the design shown in [23] is employed (Figure 4). It contains 26 slices with 2 different heights placed alternatively. The difference between the interior layer and the upper layer is 1 cm, this value has been chose being higher than the error in both devices and the minimal height possible to avoid some occlusions since in both devices the emission and the receptor are not placed in the same position.

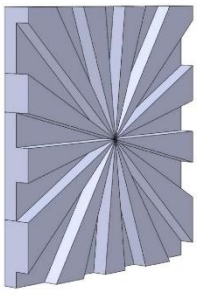


Fig. 4. Artifact to compare the spatial resolution.

4) Radiometric completeness of the object : To create a 3D point cloud by means of an RGB-D sensor, the mapping between the elements of the depth camera and those of the color camera is necessary. Although the registration between the two cameras is provided by the manufacturer, this registration has errors, which implies the quality is affected by the distance between the sensor and the objects. In order to compute these radiometric misalignments, many scenes were set up, composed of objects with the same, uniform color located over a white background. Initially, a scene segmentation is computed which provides a separation between the foreground objects and the background. Then, the color misalignment is computed through the inverse ratio between the total points belonging to the object and those points that have been assigned a wrong color after the mapping with the RGB frame, including both foreground points assigned to background colors and vice versa.

E. Comparison in terms of 3D object recognition

In addition to the metrological comparison, we want to compare the two generations of RGB-D sensors also in terms of computer vision and robotic perception applications. In particular, we have focused on 3D object recognition and pose estimation. Specifically, the task is to recognize rigid objects in clutter and occluded scenes, and estimate their position and orientation in the real world. To this end, several state-of-the-art 3D shape descriptors aimed at object recognition have been reviewed, which are publicly available that includes the algorithms from Point Cloud Library (PCL) [24]. In particular, the object recognition pipeline presented in [1] has been used. This pipeline (Figure 5) relies on global descriptors that require the notion of objects and hence deploy a specific preprocessing step based on segmentation.

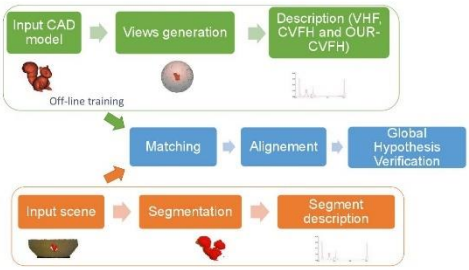


Fig. 5. Workflow for object recognition.

Since the dataset is composed of CAD models, to use this kind of models in the recognition pipeline an offline training stage is required. In this stage, a virtual camera that simulates the output of a depth sensor is uniformly placed over each vertex of a regular polygon centered in the object centroid. The overall number of virtual cameras depends on the number of vertex of the selected regular polygon. For each view, a simulated point cloud with RGB information of the model is rendered and stored in a FLANN (Fast Library for

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/JSEN.2015.2459139

Approximate Nearest Neighbors) randomized kd-forest [25] for the successive online stage. Analogously to [1] , we have used 80 views and a resolution of 150 x 150 points per view.

The deployed object recognition pipeline relies on computing a single 3D description (i.e. global) of the entire object. The first step is the segmentation of the scene, so to obtain one segment for each object. Given the specific scenario, we use the dominant plane assumption in order to fit and remove the white board that supports the objects. Then, we cluster the remaining points using a region growing method based on the Euclidean distance.

To account for different point densities between the model views and the scene, which could critically affect the performance of the descriptor matching stage, we uniformly sample all model views and scenes to normalize their resolution. Successively, as mentioned, we compute a global 3D descriptor for each scene segment as well as for each model view. In our experiments, we have tested with the state-of-the-art in 3D global description. In particular, we have compared three methods:Oriented, Unique and Repeatable Clustered Viewpoint Feature Histogram (OUR-CVHF) [26], Clustered Viewpoint Feature Histogram (CVFH) [27] and Viewpoint Feature Histogram (VFH) [28].

After that, each scene descriptor is matched to the database of model views' descriptors using fast indexing schemes - again, the randomized kd-forest implementation included in the FLANN library [25], retaining a certain number of nearest neighbors for each descriptor. Finally, to obtain the estimated 6-DOF pose for each of such nearest neighbor candidates, the pipelines using the VFH and the CVFH descriptors are paired with the Camera Roll Histogram approach presented in [27]. Conversely, in the case of the OUR-CVHF, there is no need to use such method, since OUR-CVFH already incorporates a method to estimate the 6-DOF pose implicitly from the description matching stage [26].

To improve the results, two additional stages are carried out. As for the first one, the estimated 6-DOF pose for each model instance found in the scene is refined by applying the ICP approach [17]. As for the second one, the list of model candidates associated to each scene segment which was previously obtained via Nearest-Neighbor matching is re-ordered by ranking each candidate based on the number of point inliers shared with the corresponding scene segment. This re-ranking, which is possible thanks to the fact that the pose of each model candidate is now refined via ICP, allows to weight the degree of match between each model candidate and the scene segment. This matching is not based on the similarity in the descriptor space, but on the fitting of the two surfaces, which tends to bring in higher robustness. The final step, usually referred in literature as Global Hypothesis Verification [29], is carried out to improve the results by discarding possible false positives arising from segmentation artifacts recognized as object instances.

III. Experimental evaluation

In this Section we present all the results from the experiments setup and the methodology developed in the previous sections (II-D and II-E).

A. Accuracy Results

1) Normality Results: First, the normality assumption of the dataset was tested graphically with a QQ-plot (Figure 6) and reinforced with the Anderson-Darling statistical test (Table III) the significance level employed for the critical value is the default 5%.

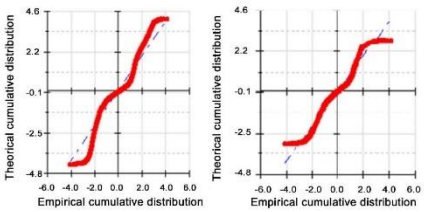


Fig. 6. QQ-Plot for Asus Xtion Pro (Left) and Kinect II (Right).

TABLE III
Results of Anderson-Darling Normality Test.

Normality Test	Statistic	Critical value	P-value	Result
Asus Xtion Pro	1198.00	0.75	$< 10^{-6}$	Non-Normal
Kinect II	574.68	0.75	$< 10^{-6}$	Non-Normal

It is clear for both approaches (graphical and numerical) that the non-normality assumption based on the presence of outliers and thus the need for non-parametric statistical methods based on robust estimators.

2) Statistical accuracy: In this section the statistical accuracy of RGB-D sensors is analysed following a two-fold process (Tables IV and V): (i) firstly, the non-parametric bootstrap sampling is applied for assessing the registration accuracy for both RGB-D sensors; (ii) secondly, the robust estimators, median and $BWMV$, are computed for those 50 scenes simulated with the dataset proposed. The results for the calibration scene demonstrate a better accuracy for the Kinect II with about 15% and visually perceptible noise reduction, that can be checked in (Figure 7). The higher accuracy yielded by Kinect II is confirmed by the results shown in Table V, providing better median and $BWMV$ values. A color map of the discrepancies for both devices is outlined in Figure 7.

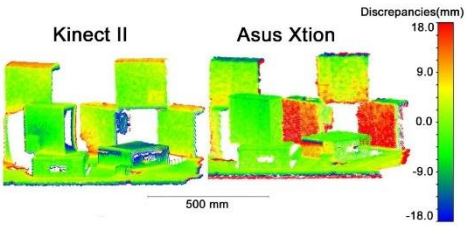


Fig. 7. Map of deviations.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/JSEN.2015.2459139

TABLE IV
RESULTS FOR ASUS XTION PRO WITH SAMPLE SIZE (N= 1000) AND BOOTSTRAP ITERATIONS (B=1500) IN MM.

Accuracy measures	Bootstrap error	Determination	Percentile bootstrap confidence interval
Median	2.168	0.115	1.932 to 2.371
SQRT Biweight Mid-variance	0.148	0.052	0.128 to 0.165
25% quantile (Q25%)	0.899	0.047	0.816 to 0.996
75% quantile (Q75%)	5.107	0.329	4.435 to 5.721
25-75% interquantile (IQR)	4.208	0.315	3.548 to 4.801

TABLE V
RESULTS FOR KINECT II WITH SAMPLE SIZE (N=1000) AND BOOTSTRAP ITERATIONS (B=1500) IN MM.

Accuracy measures	Bootstrap error	Determination	Percentile bootstrap confidence interval
Median	1.886	0.086	1.722 to 2.065
SQRT Biweight Mid-variance	0.084	0.028	0.075 to 0.093
25% quantile (Q25%)	0.854	0.047	0.758 to 0.947
75% quantile (Q75%)	3.622	0.135	3.317 to 3.886
25-75% interquantile(IQR)	2.768	0.129	2.475 to 2.995

Regarding the 3D reconstruction accuracy for the 50 scenes (Figures 8 and 9) the higher quality of the Kinect II is not demonstrated, due mainly to the reflections appearing on the board. This is a known issue [30] affecting ToF sensors since the reflections in the corners cause deformation in the acquired scene. As a consequence, the Asus Xtion Pro demonstrates higher accuracy in some scenes characterized by the presence of a small number of objects. Since the aim for the assessment is the replication of real working conditions for object recognition, the effects of reflections between the objects and the board cannot be dismissed from our analysis by, e.g., removing the board.

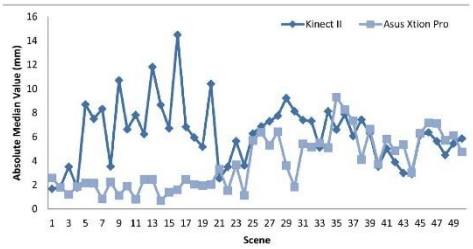


Fig. 8. Absolute Median comparison.

For the scenes with just one object (1-24) the difference is higher than the scenes having more than one object (25-44), whereas in the scenes characterized by mixed object categories (45-50) the trend is inverted, with the Kinect II providing better results. This is because with more objects, the specularities and reflections due to the board tend to be reduced, hence less deteriorating the performance of the ToF-based sensor.

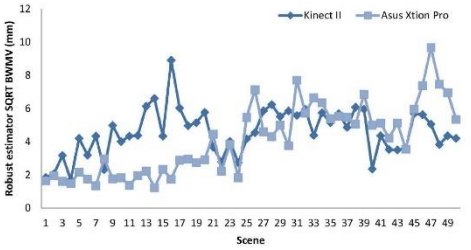


Fig. 9. Robust estimator BWMV comparison .

3) Spatial completeness: In order to assess the spatial completeness of both RGB-D sensors under real conditions, we compute the number of fitted points in the objects for all the scenes. Figure 10 shows the results for the 50 scenes setup. To compute the fitted points in each scene, we count only the numbers of points in the objects that we want to use in the scene.

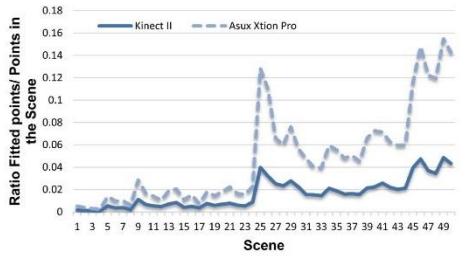


Fig. 10. Spatial completeness: fitted points for Asus Xtion Pro and Kinect II.

In this case the Asus Xtion Pro is capable to provide more points, but this factor should not be taken into account as a decision criterion as explained in section II-D3. The results for the resolution comparison are presented in Figure 11. The setup of this experiment is performed for the higher spatial resolution available, to do this, the distance between the artifact and the devices is the minimum acquisition distance, in both cases 0.5 meter. Also, it is important to remark that the angles between the artifact and the devices were minimized to avoid possible occlusions.

To compute the results, the bottom layers from both devices are extracted and compared with the ideal one. A simple visual inspection confirms the higher resolution of Kinect II (Figure 11). To calculate the specific spatial resolution, firstly is calculated the circle that fits the nearest points to the center, and secondly the shortest distance from the center to the acquired points is obtained. This radial distance is convert to the chord of the slice with more points nearest to the center. This is the real resolution of both devices, because is the minimal distance that both devices can achieve. In the case of the Asus Xtion Pro the radial distance obtained was 26.71

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/JSEN.2015.2459139

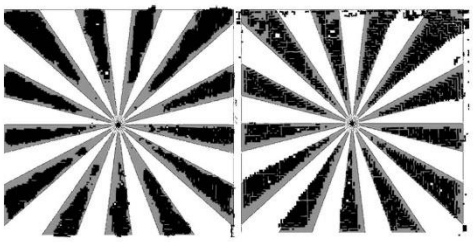


Fig. 11. Resolution results for Asus Xtion Pro (left) and Kinect II (right).

mm, while for Kinect II the minimum radial distance was 19.17 mm. For the Asus Xtion Pro the result is 6.44 mm and 4.59 mm for the Kinect II.

4) Radiometric completeness: In this section the results for the radiometric completeness are presented. Table VI shows the ratio between the wrong colored points (points out) and the fitted points (points in). To compare the behavior of the sensors, different angles and distances were tested. In particular, 3 distances (1.5 m, 0.95 m and 0.65 m) and 2 angles (90° and 45°) were chosen. As for the first position, the sensor is horizontal with respect to the objects (90°), whereas the second position (denoted in Table VI with 'up') is characterized by an angle of 45° between the sensor and the objects. As the results show, the radiometric performance is higher for Kinect II than Asus Xtion Pro since the percentage (O/I) of wrong colored points is better for Kinect II for all cases except for the 0.65 m distance. Furthermore, the variation of the radiometric completeness ratio with different distances is lower for Kinect II than Asus Xtion Pro. The Kinect II achieved roughly the same results, while the Asus Xtion Pro increased its ratio with the distance.

TABLE VI
RADIOMETRIC COMPLETENESS ANALYSIS FOR ASUS XTION PRO AND KINECT II.

	Asus Xtion Pro (Points)			Kinect II (Points)		
Position(m)	Out	In	O/I (%)	Out	In	O/I (%)
0.65	2791	20681	13.5	2427	12325	19.6
0.65 up	4172	18866	22.1	2057	10133	20.3
0.95	3037	13396	22.6	1298	6928	18.7
0.95 up	3199	14088	22.7	1704	7802	21.8
1.5	2156	6686	32.2	694	3099	22.3
1.5 up	1760	6586	26.7	599	2969	20.1

B. Object Recognition results

1) Object recognition rate: As anticipated, we present the comparison between both devices in terms of 3D object recognition and 6-DOF pose estimation for real working conditions. The pipeline presented in Subsection II-E is applied to the dataset presented with this paper, so to obtain the performance in terms of recognition rate. Since 80 views are generated from each of the 24 models, there are 1,920 possible candidates for each segment in the scene (database in Figure 2). The dataset includes 50 scenes, for a total of 126 different objects (the result of all the segmented objects in the scenes of the dataset). To carry out the descriptor matching stage between model views and scene segments, both L1 and L2 metrics were tested: we chose the former due to the higher rate experimentally demonstrated for this dataset. In Figure 12 the results for the accumulated recognition rate at different values of model candidates are presented. The accumulated recognition rate is the number of recognized objects over the total objects (126 possibilities). This criteria has been evaluated for a maximum number of nearest neighbors, denoted as 'rank'. The 'rank' goes up to 14 instances, since for higher values, the recognition rate is stabilized.

Figure 13 shows that by using a dataset acquired with the Kinect II a higher recognition rate can overall be yielded with respect to the dataset acquired with the Asus Xtion Pro, although the difference between both sensors is small. Comparing the results in terms of object categories, it can be noted that two specific categories, i.e. basic shapes and textured objects, yield worse performance with respect to the other categories. In particular, the recognition pipeline performs badly when objects at different scales are present in the scene. This is intuitively expected, since no stage of the deployed pipeline holds scale invariance with respect to the object model. As for the textured objects, we address the lower performance to the more similar geometry of the objects included in such category with respect to those belonging to the other categories. In addition, Figures 13 and 14 report the results on two smaller versions of the original dataset, respectively where the basic shapes (Figure 13) and the basic shapes plus textured objects were removed (Figure 14). Both Figures demonstrate that the recognition rate rises significantly with respect to that shown in Figure 12, whereas the difference between the two sensors still remain comparable to that depicted in the case of the complete dataset.

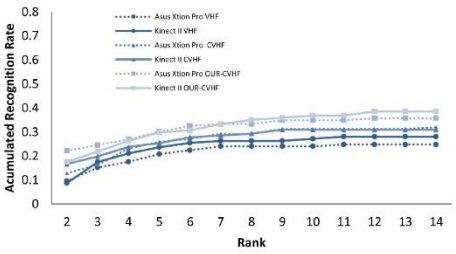


Fig. 12. Global recognition results.

2) Comparison in terms of 6-DOF pose estimation: To carry out an evaluation in terms of accuracy of estimated 6-DOF pose, we performed a comparison between the pose estimated using the object recognition pipeline and that gathered from the metrological arm (ground truth). It should be noted that in the case of highly symmetrical objects (e.g. a sphere) it is not possible to perform such evaluation in terms of rotation accuracy. For those objects presenting such symmetrical shape, only the 3-DOF pose (i.e., in terms of 3D translation) is

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/JSEN.2015.2459139

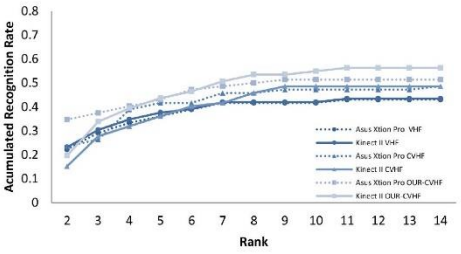


Fig. 13. Global recognition results without basic shapes.

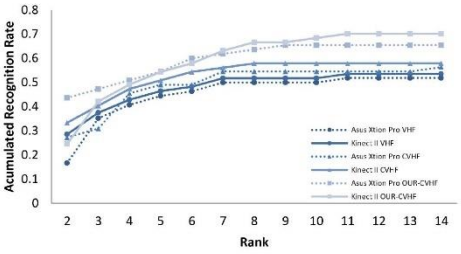


Fig. 14. Global recognition results without basic shapes and textured objects.

evaluated.

Table VII shows the results for the 6-DOF pose estimation, in terms of mean error, calculated in terms of absolute values, for each of the 6 degrees of freedom respectively. For each sensor in this table, the pose error associated to all the correctly recognized objects for the three evaluated 3D descriptors (VHF, CVHF, OUR-CVHF) was averaged together. As we can see the trend reported in the previous object recognition experiment is confirmed also in this case, the Kinect II resulting in a slightly higher accuracy with respect to the Asus Xtion Pro both in terms of translation as well as rotation error.

TABLE VII
MEAN OF ALL OBJECTS RESULTS IN THE DATASET.

	Asus Xtion Pro	Kinect II
x (mm)	6.79	5.80
y (mm)	6.29	5.84
z (mm)	5.04	4.66
Roll (deg.)	1.86	1.75
Pitch (deg.)	2.57	1.87
Yaw (deg.)	1.71	2.01

IV. Concluding remarks

Through this article a detailed comparison between the two generations of RGB-D sensors has been carried out, evaluating together two popular sensors such as the Asus Xtion Pro and the Microsoft Kinect II. More specifically, a dataset with ground truth provided by a metrological arm has been specifically acquired so to quantitatively compare the specific characteristics of both devices. This evaluation allowed us to deduce some guidelines to be able to choose the best device depending on the specific application requirements. Overall, the Kinect II has yielded better results in the main compared characteristics, although, as we demonstrated in the statistical accuracy assessment, it can be severely affected by the presence of reflections and specularities [30]. On the other hand, the Asus Xtion Pro provides higher number of fitted points but this higher value does not mean better results in the object recognition framework. This is achieve thanks to the higher resolution, which allows to the Kinect II with less accuracy in some scenes and less number of points to get better results; so the assumption exposed in section II-D3 is verified. The Kinect II also provides less error in the mapping between the depth and color frames, and it is more constant with distance variations. It is also interesting to remark, that the Kinect II is capable to work in outdoor condition of which is a feature thath the Asus Xtion Pro is not capable. As for the evaluation in terms of object recognition, the Kinect II presents a slightly higher accuracy in terms of pose estimation and higher object recognition rate. Finally it should be note that the kinect II has better results, this results can be affected in real scenes by the enviroment, but even with these issues the Kinect Ii is capable of get better results in the computer vision aplications.

Acknowledgements

Authors would like to give thanks to the Ministerio de Economa y Competitividad for the financial support given through the project (IPT-2012-1121-370000). The authors wish to thanks ITP (Industria de Turbo Propulsores) and EA (Ejército del Aire) for all advices and contribution.

References

[1] A. Aldoma, Z.-C. Marton, F. Tombari, W. Wohlkinger, C. Potthast, B. Zeisl, R. B. Rusu, S. Gedikli, and M. Vincze, “Point cloud library,” *IEEE Robotics & Automation Magazine*, vol. 1070, no. 9932/12, 2012.

[2] J. Sturm, N. Engelhard, F. Endres, W. Burgard, and D. Cremers, “A benchmark for the evaluation of rgb-d slam systems,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2012*. IEEE, 2012, pp. 573–580.

[3] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon, “Kinect-fusion: Real-time dense surface mapping and tracking,” in *10th IEEE international symposium on Mixed and augmented reality (ISMAR), 2011*. IEEE, 2011, pp. 127–136.

[4] A. Karpathy, S. Miller, and L. Fei-Fei, “Object discovery in 3d scenes via shape analysis,” in *IEEE International Conference on Robotics and Automation (ICRA), 2013*. IEEE, 2013, pp. 2088–2095.

[5] C. V. Nguyen, S. Izadi, and D. Lovell, “Modeling kinect sensor noise for improved 3d reconstruction and tracking,” in *Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012*. IEEE, 2012, pp. 524–530.

[6] K. Lai, L. Bo, X. Ren, and D. Fox, “A large-scale hierarchical multi-view rgb-d object dataset,” in *IEEE International Conference on Robotics and Automation (ICRA), 2011*. IEEE, 2011, pp. 1817–1824.

[7] “3d object recognition dataset,” 2015. [Online]. Available: http://tidop.usal.es/dataset/Objectrecognition.zip

[8] H. Gonzalez-Jorge, B. Riveiro, E. Vazquez-Fernandez, J. Martínez-Sánchez, and P. Arias, “Metrological evaluation of microsoft kinect and asus xtion sensors,” *Measurement*, vol. 46, no. 6, pp. 1800–1806, 2013.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/JSEN.2015.2459139

[9] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, pp. 1437–1454, 2012.
[10] H. Haggag, M. Hossny, D. Filippidis, D. Creighton, S. Nahavandi, and V. Puri, "Measuring depth accuracy in rgbd cameras," in *2013, 7th International Conference on Signal Processing and Communication Systems (ICSPCS)*, 2013.
[11] E. Lachat, H. Macher, M. Mittet, T. Landes, and P. Grussenmeyer, "First experiences with kinect v2 sensor for close range 3d modelling," *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 1, pp. 93–100, 2015.
[12] T. Oggier, M. Lehmann, R. Kaufmann, M. Schweizer, M. Richter, P. Metzler, G. Lang, F. Lustenberger, and N. Blanc, "An all-solid-state optical range camera for 3d real-time imaging with sub-centimeter depth resolution (swissranger)," in *Optical Systems Design*. International Society for Optics and Photonics, 2004, pp. 534–545.
[13] M. Frank, M. Plaue, H. Rapp, U. Köthe, B. Jähne, and F. A. Hamprecht, "Theoretical and experimental error analysis of continuous-wave time-of-flight range cameras," *Optical Engineering*, vol. 48, no. 1, pp. 013 602–013 602, 2009.
[14] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (tof) cameras: a survey," *IEEE Sensors Journal*, vol. 11, no. 9, pp. 1917–1926, 2011.
[15] S. B. Gokturk, H. Yalcin, and C. Bamji, "A time-of-flight depth sensor-system description, issues and solutions," in *Conference on Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04.* IEEE, 2004, pp. 35–35.
[16] "Bq witbox," http://www.bq.com/es/productos/witbox.html, 2014, 9 February 2015.
[17] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Robotics-DL tentative*. International Society for Optics and Photonics, 1992, pp. 586–606.
[18] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes," *The annals of mathematical statistics*, pp. 193–212, 1952.
[19] R. B. DAgostino, "Tests for the normal distribution," *Goodness-of-fit techniques*, pp. 367–419, 1986.
[20] H.-J. Chang, K.-C. Huang, and C.-H. Wu, "Determination of sample size in using central limit theorem for weibull distribution," *International journal of information and management sciences*, vol. 17, no. 3, pp. 31–46, 2006.
[21] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. CRC press, 1994.
[22] P. S. Horn, "Introduction to robust estimation and hypothesis testing," *Technometrics*, vol. 40, no. 1, pp. 77–78, 1998.
[23] W. Boehler, M. B. Vicent, and A. Marbs, "Investigating laser scanner accuracy," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 34, no. Part 5, pp. 696–701, 2003.
[24] R. B. Rusu and S. Cousins, "3d is here: Point cloud library (pcl)," in *IEEE International Conference on Robotics and Automation (ICRA), 2011*. IEEE, 2011, pp. 1–4.
[25] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration." *VISAPP (1)*, vol. 2, 2009.
[26] A. Aldoma, F. Tombari, R. B. Rusu, and M. Vincze, *OUR-CVFH–Oriented, Unique and Repeatable Clustered Viewpoint Feature Histogram for Object Recognition and 6DOF Pose Estimation*. Springer, 2012.
[27] A. Aldoma, M. Vincze, N. Blodow, D. Gossow, S. Gedikli, R. B. Rusu, and G. Bradski, "Cad-model recognition and 6dof pose estimation using 3d cues," in *IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011*. IEEE, 2011, pp. 585–592.
[28] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, "Fast 3d recognition and pose using the viewpoint feature histogram," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2010*. IEEE, 2010, pp. 2155–2162.
[29] A. Aldoma, F. Tombari, L. Di Stefano, and M. Vincze, "A global hypotheses verification method for 3d object recognition," in *Computer Vision–ECCV 2012*. Springer, 2012, pp. 511–524.
[30] S. Guomundsson, H. Aanaes, and R. Larsen, "Environmental effects on measurement uncertainties of time-of-flight cameras," in *International Symposium on Signals, Circuits and Systems (ISSCS07),*, vol. 1. IEEE, 2007, pp. 1–4.

4.2. Aplicación de visión artificial en registro de imágenes de varias modalidades

Esta segunda sección del capítulo contiene el artículo “*Feature matching evaluation for multimodal correspondence*” publicado en la revista de alto impacto “*ISPRS Journal of Photogrammetry and Remote Sensing*” en Mayo de 2017.

4.2.1. Resumen

Este artículo propone un estudio y evaluación de diferentes metodologías creadas para realizar un registro de imágenes con diferentes modalidades. Definimos modalidades según el espectro electromagnético en el que se generan. Por ejemplo, (RGB) para el espectro visible, IRT para imágenes térmicas correspondientes al espectro infrarrojo lejano, de intensidad o de rango para imágenes láser o de profundidad en el espectro infrarrojo cercano o medio, dependiendo del sensor. El artículo realiza una revisión de las metodologías utilizadas para realizar la comparación en este contexto específico. Finalmente se presenta un nuevo algoritmo para la mejora en el registro de imágenes. Lo primero que se realiza en este artículo es la creación de un nuevo “dataset” constituido por imágenes multimodales. La creación de este “dataset” se ha debido a las limitaciones de los “datasets” públicos actuales. El mismo incluye las siguientes modalidades: imágenes termográficas, imágenes de intensidad, imágenes visibles (RBG) e imágenes de rango. Como segundo paso se comparan diferentes algoritmos pertenecientes al estado del arte para la detección de puntos característicos (features detectors) y la correspondiente descripción de estos. El registro por plantillas o template matching es una técnica comúnmente utilizada para encontrar zonas de correspondencia entre imágenes de diferentes modalidades, y en nuestro caso las hemos adaptado para crear un método capaz de resolver el registro de imágenes por características (feature matching) y poder compararlos con los métodos clásicos de registro por puntos característicos. En total se comparan 28 diferentes combinaciones de detectores de características y descriptores, evaluando su repetibilidad y su calidad a la hora de completar el registro. Para ello, se han representado los resultados

utilizando la Característica Operativa del Receptor o (ROC- Receiving Operating Characteristic), que es una representación gráfica de la sensibilidad frente a la especificidad para un sistema clasificador binario según se varía el umbral de discriminación. Esta curva se presenta para las 28 combinaciones de detector-descriptor presentadas, resaltando las que han obtenido un mejor resultado. Finalmente, un algoritmo llamado Adaptative Pairwise Matching (APM) es creado para mejorar la robustez del proceso de registro utilizando la eliminación de "outliers" (puntos con malas características) y se hace una comparación del algoritmo introducido con otros ampliamente utilizados.

ISPRS Journal of Photogrammetry and Remote Sensing 129 (2017) 179–188

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Feature matching evaluation for multimodal correspondence

M. Gesto-Diaz [a,*], F. Tombari [b,c], D. Gonzalez-Aguilera [a], L. Lopez-Fernandez [a], P. Rodriguez-Gonzalvez [a]

[a] Cartographic and Land Engineering Department, University of Salamanca, Hornos Caleros 50, 05003 Avila, Spain
[b] DISI, University of Bologna, V.le del Risorgimento 2, Bologna, Italy
[c] CAMP, Technische Universität München (TUM), Boltzmannstr. 3, Garching b. München, Germany

ARTICLE INFO

Article history:
Received 13 August 2016
Received in revised form 4 May 2017
Accepted 8 May 2017
Available online 13 May 2017

Keywords:
Features
Keypoints
Multimodal
Image matching
Detector
Registration
Descriptor

ABSTRACT

This paper proposes a study and evaluation of approaches aimed at image matching under different modalities, together with a survey of methodologies used for performance comparison in this specific context, and, finally, a novel algorithm for image matching. First, a new dataset is introduced to overcome the limitations of existing datasets, which includes modalities such as visible, thermal, intensity and depth images. This dataset is used to compare the state of the art of feature detectors and descriptors. Template matching techniques commonly used to carry out multimodal correspondence are also adapted and compared therein. In total, 28 different combinations of detectors and descriptors are evaluated. In addition, the detectors' repeatability and the assessment of matching results based on Receiving Operating Characteristic (ROC) curve associated to all tested detector-descriptor combinations are presented, highlighting the best performing pairs. Finally, a novel Adaptive Pairwise Matching (APM) algorithm created to improve the robustness of matching towards outliers is also proposed and tested within our evaluation framework.

1. Introduction

Determining similarity between visual data is necessary in many computer vision tasks (Viola and Jones, 2001; Belongie et al., 2002; Tissainayagam and Suter, 2005; Zitova and Flusser, 2003). Methods for performing these tasks are usually based on representing an image using some global or local image properties (*features*) and comparing them using a similarity measure. However, most of the existing methods are designed for matching images within the same modality or under similar imaging conditions. They often fail when are applied to data acquired from different sensor modalities or under different photometric conditions. In such cases the sought pattern may exhibit linear or non-linear variations in the tone mapping due to changes in illumination conditions, intrinsic camera parameters, viewing positions, different modalities, etc.

The majority of matching strategies of image pairs follow a methodology that has been well introduced in Zitova and Flusser (2003). This methodology encloses four steps: (i) feature detection, (ii) feature matching, (iii) transform model estimation and (iv) image resampling.

In this paper, combinations of different state-of-the-art detectors and descriptors are analysed to find which setup gives the best performance in matching pairs of images which exhibit strong tone mapping variations due to the aforementioned reasons. In addition, an adaptive pairwise matching (APM) approach is proposed, aimed at outlier rejection to refine the transformation estimation. To carry out this evaluation, a specific dataset is introduced, which includes relevant application-wise combinations of different modalities, such as depth data (acquired with a gaming sensor, Kinect II) paired with thermal images. In the case of Kinect II, a novel approach which registers directly depth to visible can avoid the registration errors (Gesto-Diaz et al., 2015) presented in the device between colour and depth and also allows to register the Kinect device with multiple devices with different modalities, such as thermal spectrum.

There have been different works on multimodal correspondence based on self-similarity. In Huang et al. (2011) different methods to build the self-similarity descriptor are compared and applied in multimodal images (visible against LiDAR and visible with different illumination conditions). In Bodensteiner et al. (2010) a comparison to match local patch regions using descriptors prone to multimodal image matching (MI and self-similarity) is

* Corresponding author.
E-mail addresses: mgesto@usal.es (M. Gesto-Diaz), federico.tombari@unibo.it (F. Tombari), daguilera@usal.es (D. Gonzalez-Aguilera), luisloez89@usal.es (L. Lopez-Fernandez), pablorgsf@usal.es (P. Rodriguez-Gonzalvez).

http://dx.doi.org/10.1016/j.isprsjprs.2017.05.007
0924-2716/

applied. The modalities in this case were visible against infrared or LiDAR. Heinrich et al. (2012) presented a new descriptor for matching images with different modalities based on the principles of self-similarity applied to medical imagery modalities. There are also other studies presenting new breakthroughs to tackle multimodal matching. For instance, in Kim et al. (2014) authors propose a new descriptor based on the frequency of self-similarity for matching near-infrared and visible images. In Senthilnath et al. (2013) a new feature matching descriptor, Discrete Particle Swarm Optimization (DPSO), is introduced and combined with one keypoint detector. In Tombari and Di Stefano (2014) proposed a new keypoint detector based on self-dissimilarity to find interest points, applying this methodology also on a multimodal dataset. Another contribution based on SIFT (Cheung and Hamarneh, 2007) introduces a new descriptor to match across medical images. Other studies (Torabi et al., 2011), perform a comparison of several descriptors used for multimodal matching (visible and thermal), but they were paired to only one specific feature detector. In Senthilnath and Prasad (2014) an interesting variation of the framework for matching multimodal images based on SIFT and a genetic algorithm for matching is presented.

On the other hand, several works have presented registration alternatives to feature-based matching, most of them employed in medical imaging. The idea of these approaches is to use some kind of similarity among images, one of these approaches is Mutual information (MI) developed by Viola and Wells (1997). When MI became popular some novel approaches were inspired, such as the approach developed by Wachowiak et al. (2004) where a method to register images based on the normalization of the MI was presented. The approach presented by Hel-Or et al. (2014) is a method for pattern recognition in images with different modalities, with an inspiration on MI but with a different approach, called Multi Tone Mapping (MTM).

To the best of our knowledge, there is no work in literature which takes into consideration a range of modalities (visible, thermal, LiDAR intensity and depth images).

The detectors used in this work are the following ones: the well established SIFT (Scale Invariant Feature Transform) (Lowe, 1999) and SURF (Speeded-Up Robust Features) (Bay et al., 2006), and more recent approaches such as ORB (Oriented FAST (Features from Accelerated Segment Test) (Rosten and Drummond, 2006) and Rotated BRIEF (Binary Robust Independent Elementary Features) (Calonder et al., 2010; Rublee et al., 2011) MSD (Maximal Self-Dissimilarities) (Tombari and Di Stefano, 2014). These detectors are used in combination with several descriptors. Each descriptor originally proposed together with the introduced detectors is used. Except for the case of MSD that was proposed without any specific descriptor. Some descriptors also are included in this work that have already been used for multimodal correspondence in previous works, e.g., LSS (Local Self Similarity) (Shechtman and Irani, 2007) and HOG (Histogram Oriented Gradients) (Dalal and Triggs, 2005). Furthermore, MI (Mutual information) (Viola and Wells, 1997) and MTM (Multi-Tone Mapping) (Hel-Or et al., 2014) are two template matching solutions extensively used for multimodal correspondence. To include them into our evaluation framework these two popular techniques are adapted to work like a descriptor. The 4 detectors (MSD, ORB, SIFT and SURF) combined with the 7 descriptors (HOG, LSS, MI, MTM, ORB, SIFT and SURF) provide 28 possible combinations for the comparison.

Finally, a novel Adaptive Pairwise Matching (APM) for pairwise image matching is proposed. This method automatically selects the best correspondences to determine the transformation estimation between a pair of images, including an outlier removal method, that can be RANSAC (Random Sample Consensus) (Fischler and Bolles, 1981) or LMedS (Least Median of Squares) (Zhang et al., 1995).

Importantly, public datasets for quantitative evaluation of algorithms for multimodal correspondence are quite limited. For this reason, we start by adopting the methodology and structure of the dataset presented in Mikolajczyk and Schmid (2005) (hereinafter referred to as Oxford dataset), which represents a reference benchmark for pairwise image matching, but only contains pairs of images acquired with an optical camera. Then, we add several image pairs acquired under different modalities, hence obtaining a bigger dataset, which we plan to publicly release upon publication of this work.[1] For what concerns the performance evaluation, we present comparative results in terms of keypoint repeatability for the evaluated detectors, as well as ROC curves for the evaluated descriptors, and the number of images registered using the proposal outlier rejection method compared to RANSAC and LMedS.

This paper has been structured as follows: Section 2 presents the materials used to create the image dataset for multimodal matching and the methods used to compare the detectors, descriptors and the novel matching algorithm. The experimental results are reported and discussed in Section 3, while final remarks and conclusions are drawn in Section 4.

2. Materials and methods

2.1. Multimodal dataset

Fig. 1 shows a set of 10 image pairs that we have added to those originally included in the Oxford dataset to perform our evaluation. This accounts for a total of 15 images pairs.

The modalities taken into account in this dataset are: thermal, visible, LiDAR intensity and depth images from a gaming sensor, Kinect II. The first four image pairs are visible with LiDAR. The fifth image pair is depth image with thermal. Finally, five more images pairs combining visible with LiDAR intensity images were included.

These image pairs are from two different sources: one is a synthetic, created virtually from a 3D real object with different acquired modalities (the first four image pairs). For these cases of synthetic images, it can be easily controlled rotation, scale and shear to evaluate the behaviour of the detectors and descriptors in the images affected for scale, rotation and shear. Please note, that in this case, the ground truth is perfectly defined by the transformation applied. In the remaining cases, the images have been extracted from real cases, where a method to obtain the ground truth with the transformation between each pair was used. In these cases, the sensors and positions used for acquiring the images are not the same, such as thermal and visible images; therefore, the intrinsic and extrinsic parameters are different. In addition to this, the tone mapping changes in a non-linear way due to changes in illumination conditions and its different modality. To quantitatively compare different solutions on this dataset, the ground truth represented by the transformation between each image pair needs to be obtained. The transformation used as ground truth is the fundamental matrix (Luong and Faugeras, 1996)

$$xFx'^{T} = 0 \tag{1}$$

where x and x' are vectors of matching points presented in both images expressed in homogeneous coordinates.

This matrix (Eq. (1)) provides the transformation for a set of matching points between a pair of images. The methodology for the estimation of the fundamental matrix is the same as in Mikolajczyk and Schmid (2005). The fundamental matrix between the reference image and the other image in a particular dataset is

[1] Available for reviewing at http://tidop.usal.es/dataset/datasetmultimatching.7z.

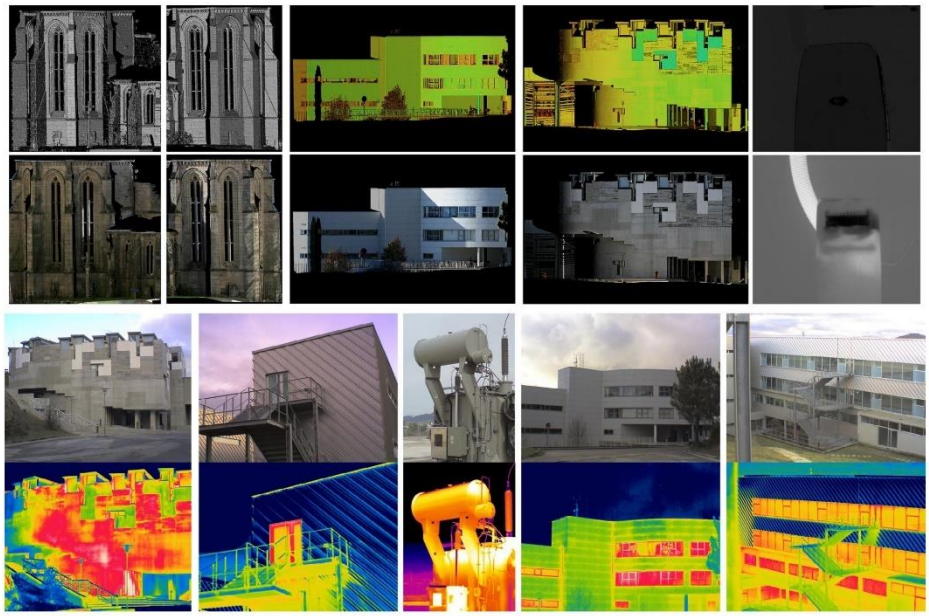


Fig. 1. Pairs of images selected for the multimodal matching dataset.

computed in two steps. First, a small number of point correspondences are selected manually between the reference and the other image. These correspondences are used to compute an approximate fundamental matrix between the images, and the other image is warped by its fundamental matrix so that it is roughly aligned with the reference image. Second, a standard robust fundamental matrix estimation algorithm (Bolles and Fischler, 1981) is used to compute an accurate residual fundamental matrix between the reference and warped image (using hundreds of automatically detected and matched interest points). The composition of these two matrices (approximate and residual) gives an accurate fundamental matrix between the reference and other image. The manual selection from the first step of the points was devoted to dismiss outliers in the fundamental matrix computation, so to obtain pixel precision. Please note that the application of a refinement procedure, in a multimodal case, could lead to higher errors that by manual procedure. By a selection of a high number of points, and their robust refinement, the final error propagation is in lower than one pixel. This precision is higher than the required for the comparison method proposed by Mikolajczyk and Schmid (2005).

2.2. Methodologies

2.2.1. Detectors and repeatability

The workflow used for the comparison is presented in Fig. 2. First, the detector is computed for each pair of images. Then, the detectors find characteristic features, represented by points with an associated size around it, called keypoints. With these keypoints in both images, the repeatability can be calculated, as the ratio between the correct keypoints and all keypoints in one image pair. To establish if a feature has been presented in both images, the method introduced by Mikolajczyk and Schmid (2005) is used. This proposal consists in comparing the overlap of the meaningful information provided for each keypoint. If there are enough overlap (in this work is the 65%) between the meaningful information of both keypoints, it is assumed that these keypoints represent the same feature in both images. To check if the same keypoint is present in both images, a transformation of one of the images needs to be computed, and then the overlap is estimated using the (Mikolajczyk and Schmid, 2005) method. This transformation is performed using the ground truth of the dataset.

2.2.2. Descriptors and ROC curves

Following the pipeline of Fig. 2, the next step consists in associating each previously retrieved keypoint with a descriptor, this allowing to create a unique and defined feature for each keypoint. Descriptors encode the local neighbourhood around the keypoint in a way that is invariant or robust to transformations such as scale, rotation or illumination changes. Afterwards, these features are matched to the most similar one in its image pair, thus yielding a set of correspondences. Finally, a ROC (Receiver Operating Characteristic) curve is built to measure the precision and recall of the retrieved correspondences.

It has been used precision and recall expressed as $precision = tp/(tp + fp)$ and $recall = tp/(tp + fn)$, where tp is the number of true positive correspondences; fp is the number of false positive correspondences and fn is the number of false negative correspondences. These values are determined by means of a threshold that spans the operating points of the curve. Specifically, this threshold comes from a coefficient, k, multiplied by the maximal matching distance from all the correspondences. Particularly, the coefficient, k, spans between 0 and 1 with a step of 0.1. For example, the first point of the curve ROC has been calculated with

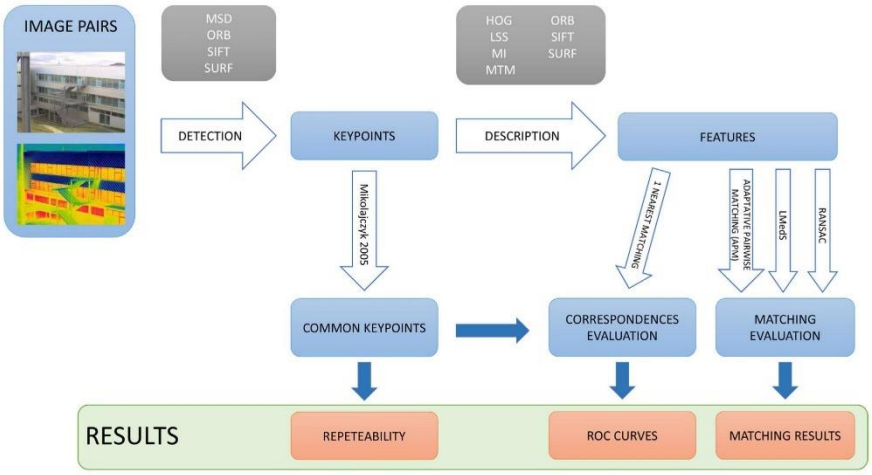


Fig. 2. Pipeline used to compare detectors, descriptors and matching methods in multi-modal image pairs.

all the correspondences with a distance lower than 0.1 of the maximal distance in the set of correspondences. Hence, precision is the ratio between the number of correct correspondences over all obtained correspondences that are over the threshold, while recall is the number of correct correspondences from all features in the image pair.

On the other hand, the adaptation of template matching (MTM and MI) to feature matching was carried out as follows. Due to the non-existence of descriptors in the case of MTM and MI, a comparison between all the information associated to each feature in an image pair was performed. For example, a feature encountered in the first image of the pair needs to be compared with all the features in second image. The correspondence is established with the match with the minimum distance between descriptors. Because of the variable sizes in features, some normalization works needed to be performed as follows:

- First, for each keypoint in one image, a patch based on a size provided by the detector is retrieved (left image in Fig. 3)
- Secondly, the patch is rotated (90° in the centre image Fig. 3) according to the orientation provided by the detector.
- Finally the patch is normalized to a common size, being 16×16 in this work(right image in Fig. 3) following a bicubic interpolation (Keys, 1981) over 4×4 pixel neighbourhood. This 16×16 value was chosen following the recommendations and the best results shown in Hel-Or et al. (2014). The MI and MTM distances are computed among all the patches in the first image with all the patches in the second image. A final set of correspondences is created with the matched features between images with the minimum distance.

2.2.3. Adaptive pairwise matching

In this section, a new algorithm for outliers removal and improvement of the matching called Adaptive Pairwise Matching (APM) is presented. The main goal of matching is to find a relationship between the image pairs, normally represented with the fundamental matrix (**F**). The idea of APM is to reduce the great number of outliers that usually the images with different modalities enclose and improve the efficiency of the matching methods for fundamental matrix estimators (e.g. the 8 Point algorithm) (Hartley and Zisserman, 2003) with different outliers removals, such as RANSAC (Bolles and Fischler, 1981) and LMedS (Zhang et al., 1995).

Algorithm 1. Adaptive Pairwise Matching(APM)

Input: Matches Image 1 to 2 (m1to2), Matches Image 2 to 1 (m2to1).
Output: transformation, inliers
procedure ADAPTATIVE MATCHING
  $maxdistanceltor \leftarrow max(m1to2(distance))$
  $maxdistancertol \leftarrow max(m2to1(distance))$
  $maxdistance \leftarrow \max(maxdistanceltor, maxdistancertol)$
  for $k := 0 < 1$ **do**
    $matchesltor \leftarrow m1to2(distance \leqslant i * maxdistance)$
    $matchesrtol \leftarrow m2to1(distance \leqslant i * maxdistance)$
    Ratio Test
    Symmetry Test
    for all *matchesltor* **do** ▷{Retrieve all common matches}
      $matches \leftarrow matches1to2 = matches2to1$
      End for;
    $inliers \leftarrow RansacTransformation(matches)$
    if $inliers \geqslant thold$ *break* **End if;**
    $k \leftarrow k + 0.1$
    End for;
  if $inliers \geqslant thold$ **then**
    {Transformation refinement}
  $Transformation \leftarrow RansacTransformation(inliers)$
else
  Transformation cannot be solved
  End if;
End procedure;

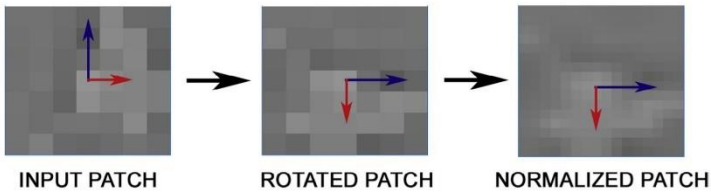


Fig. 3. Patch Normalization for MTM and MI feature matching adaptation.

The first step of the algorithm consists in matching the features in the images pair. To this end, a n-neighbour is conducted for all features, this means, for each feature in one image, a number n of the most similar features in the other image is retrieved. The number of selected neighbours is based on the shortest distance between features, and the total amount of matches is equal to the number of keypoints multiplied by n ($n = 2$ in our evaluation).

Since we are taking into account only pair-wise image matching, in order to improve the algorithm's robustness, we perform matching along both directions, i.e., using the first image as reference while searching for the best correspondences in the other image, and vice versa.

Successively to the matching stage, we iterate in order to detect outliers and reject wrong matches. Specifically, the algorithm iterates around the distance of the input matches, so the best matches can be selected. At each iteration, all matches under a certain threshold are retrieved. At each iteration, this threshold is incremented by 10% of the maximum distance found among all current matches. Consequently, in each iteration step, the number of considered matches increases. The retrieved matches at each iteration are tested to check whether they hold a consensus for a reliable transformation between the image pair. To this aim, some tests are performed at each iteration. Firstly, a ratio-distance test between pairs of correspondences is performed similarly to the one introduced by Lowe in Lowe (1999). In particular, for each extracted point the distance ratio between the two best candidates in the other image is compared with a threshold. If it is obtained a low distance ratio, the match could be more distinguishable, due to the difference of distances between them, being reduced the probability of a mismatch. According to the Probability Distribution Function, a threshold >0.8 provides a good separation among correct and incorrect matches. Greater the ratio value, greater the amount of matched points. After that, a symmetry test is performed to retrieve just the matches presented in both directions (left to right and right to left). The last test is to apply the classical methods for outliers rejection (RANSAC or LMedS) in order to calculate the transformation with the best and most reliable matches available after the tests. If this last test returns a number of inliers bigger than a threshold (thold), the iterative stage is terminated. Finally, the algorithm calculates the final transformation between both images based on the 8-point algorithm (Hartley and Zisserman, 2003) using the inliers returned for the rejection method.

One advantage of this approach is that besides RANSAC, other algorithms such as LMedS can be used to reject the outliers and can be implemented with any outlier rejection method. The advantages of the APM algorithm is to use the best and most reliable points, turning out very useful when dealing with complex multimodal images. To analyse the quality of this novel algorithm, it has been compared with RANSAC and LMedS. This quality comparison is performed by the number of pair of images that each algorithm has been able to solve from the n nearest matched features in this work $n = 2$) Fig. 2.

Finally a method for the comparison between RANSAC, LMeDS and the APM method is performed. This has been done use the method proposed in Senthilnath and Prasad (2014) where the matching correspondence accuracy is obtained by checking the number of inliers get for any of the methods compared against the correct inliers checked with the ground truth.

3. Experimental evaluation

In this section, the results for all multimodal pairs have been presented for detectors, descriptors and matching. Firstly, the

Table 1
Detectors repeatability grouped by categories.

Dataset	Detector				Dataset	Detector			
Oxford	MSD	SIFT	ORB	SURF	Visible plus intensity	MSD	SIFT	ORB	SURF
No effect	0.371	0.187	0.356	0.256	No effect	0.696	0.361	0.436	0.7046
Scale					Scale				
1.5	0.371	0.187	0.356	0.256	1.5	0.685	0.371	0.425	0.715
2	0.304	0.159	0.285	0.243	2	0.562	0.304	0.349	0.586
2.5	0.249	0.135	0.228	0.231	2.5	0.461	0.249	0.286	0.480
3	0.204	0.115	0.182	0.219	3	0.378	0.204	0.234	0.394
Rotation					Rotation				
5	0.315	0.166	0.299	0.210	5	0.651	0.358	0.425	0.701
15	0.236	0.141	0.236	0.147	15	0.488	0.269	0.319	0.525
25	0.177	0.106	0.177	0.110	25	0.366	0.201	0.239	0.394
35	0.133	0.079	0.133	0.082	35	0.274	0.151	0.179	0.295
View Point Angle Change (Shear Effect)					View Point Angle Change (Shear Effect)				
10	0.268	0.145	0.260	0.182	10	0.553	0.305	0.361	0.596
15	0.201	0.123	0.205	0.128	15	0.415	0.228	0.271	0.447
20	0.150	0.092	0.154	0.096	20	0.311	0.171	0.203	0.335
25	0.113	0.069	0.115	0.072	25	0.233	0.128	0.152	0.251
Visible plus Thermal	MSD	SIFT	ORB	SURF	Depth plus Thermal	MSD	SIFT	ORB	SURF
	0.604	0.395	0.554	0.580		0.158	0.074	0.108	0.250

results corresponding to the repeatability of the detectors grouped by modalities are presented in Table 1. The first case is for visible plus Intensity (Laser-scanner) images, the second one is Depth (Kinect II) plus Thermal images, the third one is visible plus Thermal images and finally the case with the repeatability from the Oxford dataset. The detectors are setup with the default parameters. In the case of MSD, the default parameters are always used except in the case of saliency, when the parameters are set to a threshold of 100, in order to deal with enough points in all categories. The repeatability is the number of keypoints that represents the same feature in both images over the total number of keypoints, spanning in the factor k defined in the previous section. The results are grouped for the modalities matched. For example, the first group is related to visible images with thermal images. This has been done to find out which combination of detectors and descriptors are more prone to match these modalities of images and to avoid skewing the results for all images due to the different behaviour.

3.1. Detectors and repeatability

In this section are presented the results for the detectors and its repeatability for pairs of images, using the different modalities of the four categories of the dataset (Table 1). The repeatability, as it was mentioned in the methodology, was calculated using the method proposed by Mikolajczyk and Schmid (2005). As a result, we can analyse the behaviour of the algorithms used for the first matching step, the detection of keypoints. For the case of Oxford and Visible plus intensity the results for different rotations and scale are presented. In the datasets visible plus Thermal and depth sensor plus Thermal present variations in scale, rotation and shear .

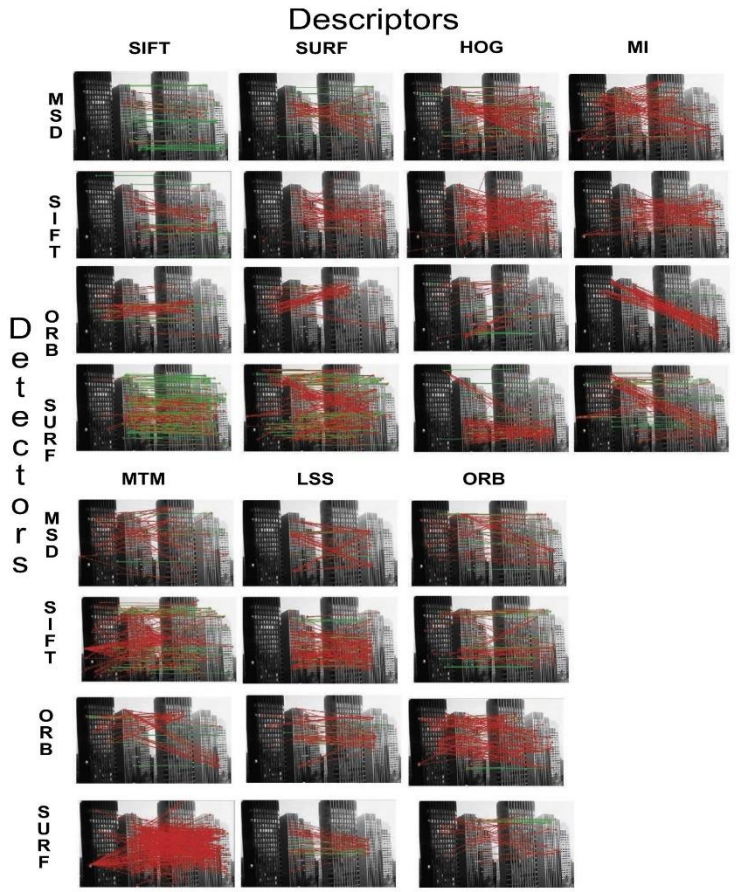


Fig. 4. Qualitative results in the form of correspondences between a pair example for all combinations. In green the right matches and in red the wrong matches according to the ground truth. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

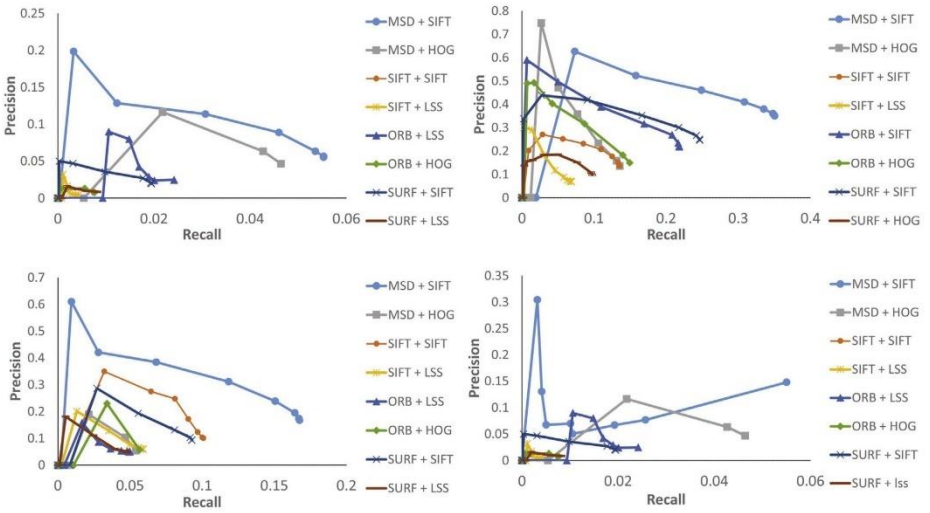


Fig. 5. ROC curves: (a) Oxford Dataset, (b) Visible plus Intensity, (c) Visible plus Thermal, (d) Depth plus Thermal.

The results show a similar behaviour for MSD and SURF, being the detectors with a higher rate of repeatability. In theory, these two detectors can provide better results for the feature matching. Furthermore, it is necessary to confirm if the retrieved keypoints represent an enough distinguishable feature. The results shows that the influence of scale or rotation is lower compared with the change of viewpoint. Due to this results for higher view of point angle change, it will be hard to compute a correct matching between these pairs of images. To do that, as anticipated in the previous section, a feature description with the 7 descriptors has been calculated for each detector, with a final number of 28 combinations. The features in one image are matched with the most similar correspondence in its pair, creating a ROC curve based on the correct and wrong matches. In Figs. 5 and 6, the results are presented grouped by categories such as in the case of detectors. It is necessary to mention that in order to present a clear graph, only the two best descriptors for each detector are presented; this makes a total of eight results for each group.

3.2. Descriptors and ROC curves

In this section is analysed which combination of detectors and descriptors provides a higher number of correct matches. An example of the correct and wrong correspondences for a pair of images with the 28 combinations can be found in Fig. 4, which shows an example of true positives, in green, and false positives, in red, and allows to see the behaviour of the combinations in different aspects, such us, spacing of features, outliers and correct matches. Fig. 4 shows that MSD and SURF detectors combined with SIFT descriptor present a great number of correct matches against wrong matches. However, this is not the case for the spacing of features since almost all matches are concentrated in a specific area. On the other hand, HOG descriptor presents better spacing in combination with MSD detector but lower rate of true versus false positives correspondences. This step shows all the correspondences presented before applying the APM algorithm. To make more visible the correct and wrong correspondences, just those correspondences with a matching-distance lower than a 40% of the worst matching-distance are outlined (Figs. 5 and 6).

Analysing the results of Figs. 5 and 6, it has become clear which group of modalities can provide better results, being visible plus intensity. This could be explained by the relationship between the wavelength of the laser scanners (inside the interval of 0.5–1.5 μm) with the visible images' spectrum. On the other hand, the results of visible plus thermal are not affected by this relationship, so the results are explained by the detectors and descriptors combinations only. The other groups present a lower response in the ROC curve, being more difficult to achieve a correct matching.

Regarding the best detector, MSD is the one offering the best ROC curve for all groups. This not only means that MSD provides a bigger amount of common points, as seen in the case of the repeatability (Table 1), also means that these points represent more distinctive features. This fact has been proven by comparing it with SURF; despite having a similar results in the category of repeatability, MSD has been able to offer a bigger amount of correct matches. In relation to the others detectors, they obtained similar results and consequently it is complicated to establish an order of quality.

Concern to matches, the combination MSD plus SIFT is the one returning the best results in almost each case. The combination MSD plus HOG has shown good results in the Oxford Dataset case. It can be deduced that the descriptors with the best results for multimodal matching are SIFT, HOG and LSS, considering that, among the 8 results (the two best descriptors for each detector), one of them is always presented in the combination with the detectors. The combination Depth plus Thermal offers the worst results. It is hard to find a suitable scene for Thermal and depth registration where a enough number of keypoints can be matched. In particular, we would need a scene with different geometries (i.e. shapes) and materials (i.e. temperatures) to become a good candidate for matching depth and Thermal images, respectively.

It has to be stressed that the use of MI and MTM as descriptors did not produce the expected results. Both cases obtained a lower performance in comparison to the classic solutions as SIFT. Besides,

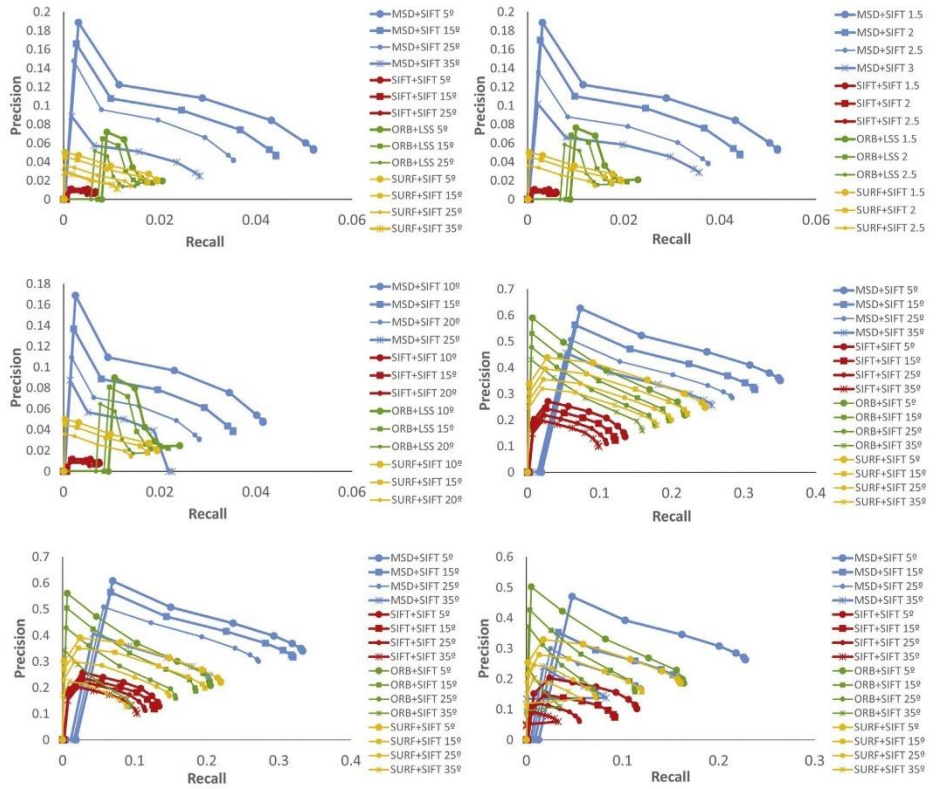


Fig. 6. ROC curves: (a) Oxford Dataset Rotation, (b) Oxford Dataset Scale, (c) Oxford Dataset View Point, (d) Visible plus Intensity Rotation, (e) Visible plus Intensity Scale, (f) Visible plus Intensity View Point.

given the fact that this solution needs to make a template matching with all the keypoints in order to conduct the match, it resulted an inefficient and slow solution. In the part of Rotation, Scale and Point of View change, we can see that the change of point of view is more significant and give lower results for both cases. Also when the change is higher the results go low. Can be analysed with more details after computing the rate of correct inliers for the transformation. Also as can be deduced that in some cases the repeatability and recall are low, being hard to accomplish the transformation, that is why the matching process in multimodal images become so hard.

3.3. Adaptive pairwise matching

The number of images pairs solved with RANSAC and LMedS is reduced compared with the number of correct pairs solved for the APM proposed algorithm. In Table 2 has been presented the combinations able to perform at least one pair.

The results have a clear correlation with the previous ones of the ROC curves, where the combinations with higher repeatability and precision have been able to match a higher number of images pairs. Also is confirmed that the APM algorithm improves the results from RANSAC and LMedS, being this new approach more prone to the multimodal image matching. It is also clear that the combination MSD plus SIFT gives the higher rank (ROC curves) in the matching stage, making this combination the best for multimodal matching purposes.

Finally, from all the correct image pairs matched, a comparison to see the rate of the correct outlier removal is presented. Using the method proposal by Senthilnath and Prasad (2014). In Table 3 the results are presented for all pairs correct matching in all the datasets present and also for the cases where a virtual modification of one image was performed. Only the combinations that are able to solve the transformation are presented, in the hard case, MSD

Table 2
Results for the matching stage(expressed in percentage considering the total of 15 image's pairs).

Combination Descriptor + Detector	APM	RANSAC	LMedS
MSD + SIFT	0.53	0.20	0.20
SURF + SIFT	0.40	0.20	0.20
MSD + HOG	0.40	0.13	0.13
SIFT + HOG	0.40	0.13	0.13
SURF + LSS	0.33	0.13	0.13
SIFT + SIFT	0.33	0.13	0.13
SURF + ORB	0.33	0.13	0.13
MSD + LSS	0.33	0.13	0.13

Table 3
The rate of the correct inliers for all the correct transformation estimations.

Dataset Oxford												
Detector plus Descriptor	APM	RANSAC	LMedS									
MSD + SIFT	0.95	0.96	0.75									
MSD + HOG	0.95	0.88	0.75									
ORB + LSS	0.90	0.80	0.75									
SURF + SIFT	0.90	0.80	–									
Scale		1.5			2			2.5			3	
	APM	RANSAC	LMedS	APM	RANSAC	LMedS	APM	RANSAC	LMedS	APM	RANSAC	LMedS
MSD + SIFT	0.98	0.92	0.80	0.94	0.93	0.70	0.90	0.80	0.95	0.95	0.80	0.65
MSD + HOG	0.90	0.90	0.75	0.91	0.90	0.72	0.89	0.81	0.95	0.95	0.78	0.75
ORB + LSS	0.88	0.93	0.81	0.88	0.90	0.71	0.88	0.78	–	0.78	0.78	–
SURF + SIFT	0.92	0.90	0.73	0.90	0.90	–	0.85	0.80	0.80	–	–	–
Rotation		5			15			25			35	
	APM	RANSAC	LMedS	APM	RANSAC	LMedS	APM	RANSAC	LMedS	APM	RANSAC	LMedS
MSD + SIFT	0.96	0.95	0.75	0.98	0.80	0.77	0.97	0.80	0.8	0.90	0.65	–
MSD + HOG	0.90	0.95	0.7	0.92	0.80	0.69	0.90	0.78	–	0.80	–	–
ORB + LSS	0.93	0.95	0.80	0.97	0.85	0.70	0.90	0.82	–	–	–	–
SURF + SIFT	0.88	0.95	0.78	0.90	0.88	0.71	0.80	–	–	–	–	–
View Point		5			10			15			20	
	APM	RANSAC	LMedS	APM	RANSAC	LMedS	APM	RANSAC	LMedS	APM	RANSAC	LMedS
MSD + SIFT	0.96	0.95	0.75	0.98	0.80	0.77	0.97	–	–	–	–	–
MSD + HOG	0.90	0.95	0.7	0.90	0.80	–	0.90	–	–	–	–	–
ORB + LSS	0.93	0.95	0.80	0.85	–	–		–	–	–	–	–
Dataset Visible plus Intensity												
Detector plus Descriptor	APM	RANSAC	LMedS									
MSD + SIFT	0.95	0.91	0.80									
MSD + HOG	0.95	0.91	0.80									
ORB + LSS	0.94	0.90	0.75									
SURF + SIFT	0.95	0.90	0.75									
Scale		1.5			2			2.5			3	
	APM	RANSAC	LMedS	APM	RANSAC	LMedS	APM	RANSAC	LMedS	APM	RANSAC	LMedS
MSD + SIFT	0.98	0.92	0.80	0.94	0.93	0.70	0.90	0.80	0.85	0.90	0.80	0.65
SURF + SIFT	0.90	0.90	0.75	0.91	0.90	0.72	0.89	0.81	0.85	0.90	0.78	0.75
ORB + SIFT	0.88	0.93	0.81	0.88	0.90	0.71	0.88	0.78	–	0.78	0.78	–
MSD + HOG	0.92	0.90	0.73	0.90	0.90	–	0.85	0.80	0.80	–	–	–
Rotation		5			15			25			35	
	APM	RANSAC	LMedS	APM	RANSAC	LMedS	APM	RANSAC	LMedS	APM	RANSAC	LMedS
MSD + SIFT 0.98	0.91	0.80	0.91	0.80	0.75	0.95	0.90	0.75	0.70	0.85	–	–
SURF + SIFT	0.90	0.88	0.87	0.85	0.95	0.77	0.95	0.80	0.75	0.80	–	–
ORB + SIFT	0.95	0.80	0.81	0.75	0.95	0.95	0.77	0.75	0.69	0.88	–	–
MSD + HOG	0.80	0.77	0.83	–	–	–	–	–	–	–	–	–
View Point		5			10			15			20	
	APM	RANSAC	LMedS	APM	RANSAC	LMedS	APM	RANSAC	LMedS	APM	RANSAC	LMedS
MSD + SIFT	1.0	0.95	0.75	0.79	0.80	–	–	–	–	–	–	–
SURF + SIFT	1.0	0.95	0.75	0.80	–	–	–	–	–	–	–	–
ORB + SIFT	0.8	–	–	–	–	–	–	–	–	–	–	–
MSD + HOG	0.75	–	–	–	–	–	–	–	–	–	–	–
Dataset Depth plus Thermal												
Detector plus Descriptor	AP	RANSAC	LMedS									
MSD + SIFT	0.95	–	–									
Dataset Visible plus Thermal												
Detector plus Descriptor	AP	RANSAC	LMedS									
MSD + SIFT	0.95	0.9	0.8									
MSD + HOG	0.80	0.85	0.7									

+ SIFT is the only one to achieve the transformation. In the cases that there are several combination solving the transformation, the best four are presented. As we can see the ratio of inliers using the method APM is always higher compared with using only RANSAC or LMedS. Also we can check that when exist some big change of scale, rotation or viewpoint the success rate decrease.

4. Conclusions

This paper presents a survey and evaluation of multimodal image matching techniques, together with a novel image matching algorithm and a novel public multimodal image dataset. The advantages of the dataset lie in its selection of modalities and in its ground truth, which allows the verification of the work and the results.

Several detector and descriptors have been tested, evaluated, and ranked according to different multimodal cases. It has been shown that the novel approach MSD offers a higher performance than the classical approaches when detecting common points in images of different modalities. The retrieved points with MSD in combination with descriptors as SIFT, HOG and LSS have been able to provide a huge number of correct matches; therefore, the

estimation of the fundamental matrix has been easily solved. It is worth to remark that the template matching techniques as MI or MTM, which provided good results for template multimodal correspondence, proved to be inefficient when working as the adaptation to feature multimodal matching. Compared with the matching between the same modalities, the results in multimodal are clearly lower, since it is really difficult to match images with different modalities and even more when there are some effect such as rotation or point of view changes. Also the modalities compared are not equally easy to solve. For example, in the case of visible against intensity the results are more easily compared than visible and Thermal. It also should be noted that in the cases where the scale, orientation and point of view are real and not simulated, the results getting worse in terms of matching precision and recall. This indicates that, these methods for multimodal matching should be used in cases where the sensors do not have high changes. Last but not least, it should be remarked that the worst case for multimodal matching is when images coming from depth and Thermal modalities. In those cases where a virtual orientation, scale and point of view change were applied, it can be noted, that the condition that more decreases the rate of the matching stage was the variation of the point of view.

The new APM has proved to provide better results than the classical matching approaches based on RASAC or LMedS. This new approach is prone for those cases where the combination detector plus descriptor gives low precision and recall, and also confirm that the best combination of detector and descriptor is MSD plus SIFT. The main contribution of the APM compared to RANSAC or LMEDS is that, the APM method offers a previous step to outlier rejection, where the more unique features and more similar in the pair are matched, increasing the quality of the matches before the outlier rejection gives higher rate of success in the matching process, as is shown in the matching correspondence accuracy. The APM strategy has been, in some case, the only outlier removal method that can solve images in some dataset, always using the detector and descriptor MSD and SIFT, respectively.

Accordingly to the global detectors, descriptors and matching repeatability analysis, it is concluded that the combination able to provide better results is MSD plus SIFT. This combination has been capable to offer more correct matches in every type of modality. Furthermore, MSD plus SIFT in combination with APM has demonstrated to give the best results in terms of multimodal correspondence. It should be noted too the higher number of inliers using our approach in comparison with the case of RANSAC and LMedS, even at 60% of the rate, this is clear in the way that it works, that filters several matches before enter in the outlier removal stage.

The future works are aimed to the enlargement of the public dataset. Also, given its public nature, an improvement is expected thanks to the participation of the scientific community in its evaluation. A promising research line, considering its great results, will be the search of a complementary descriptor for MSD detector. Add more detectors and descriptors prone to multimodal matching to use in the comparison could be interesting. More images with real scale, orientation and point of view, from two different sensors can be also useful to improve this public dataset.

References

Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf: speeded up robust features. In: Computer Vision–ECCV 2006. Springer, pp. 404–417.

Belongie, S., Malik, J., Puzicha, J., 2002. Shape matching and object recognition using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell. 24 (4), 509–522.

Bodensteiner, C., Huebner, W., Jüngling, K., Müller, J., Arens, M., 2010. Local multimodal image matching based on self-similarity. In: Image Processing (ICIP), 2010 17th IEEE International Conference on. IEEE, pp. 937–940.

Bolles, R.C., Fischler, M.A., 1981. A RANSAC-based approach to model fitting and its application to finding cylinders in range data. In: IJCAI, vol. 1981. pp. 637–643.

Calonder, M., Lepetit, V., Strecha, C., Fua, P., 2010. Brief: binary robust independent elementary features. Comput. Vis.–ECCV 2010, 778–792.

Cheung, W., Hamarneh, G., 2007. N-sift: N-dimensional scale invariant feature transform for matching medical images. In: Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on. IEEE, pp. 720–723.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1. IEEE, pp. 886–893.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24 (6), 381–395.

Gesto-Diaz, M., Tombari, F., Rodriguez-Gonzalvez, P., Gonzalez-Aguilera, D., 2015. Analysis and evaluation between the first and the second generation of rgb-d sensors. Sens. J. IEEE 15 (11), 6507–6516.

Hartley, R., Zisserman, A., 2003. Multiple View Geometry in Computer Vision. Cambridge University Press.

Heinrich, M.P., Jenkinson, M., Bhushan, M., Matin, T., Gleeson, F.V., Brady, M., Schnabel, J.A., 2012. Mind: modality independent neighbourhood descriptor for multi-modal deformable registration. Med. Image Anal. 16 (7), 1423–1435.

Hel-Or, Y., Hel-Or, H., David, E., 2014. Matching by tone mapping: photometric invariant template matching. IEEE Trans. Pattern Anal. Mach. Intell. 36 (2), 317–330.

Huang, J., You, S., Zhao, J., 2011. Multimodal image matching using self similarity. In: Applied Imagery Pattern Recognition Workshop (AIPR), 2011 IEEE. IEEE, pp. 1–6.

Keys, R.G., 1981. Cubic convolution interpolation for digital image processing. IEEE Trans. Acoust. Speech Signal Process. 29 (6), 1153–1160.

Kim, S., Ryu, S., Ham, B., Kim, J., Sohn, K., 2014. Local self-similarity frequency descriptor for multispectral feature matching. In: Image Processing (ICIP), 2014 IEEE International Conference on. IEEE, pp. 5746–5750.

Lowe, D.G., 1999. Object recognition from local scale-invariant features. Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, vol. 2. Ieee, pp. 1150–1157.

Luong, Q.-T., Faugeras, O.D., 1996. The fundamental matrix: theory, algorithms, and stability analysis. Int. J. Comput. Vis. 17 (1), 43–75.

Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. 27 (10), 1615–1630.

Rosten, E., Drummond, T., 2006. Machine learning for high-speed corner detection. In: Computer Vision–ECCV 2006. Springer, pp. 430–443.

Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. Orb: an efficient alternative to sift or surf. In: Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, pp. 2564–2571.

Senthilnath, J., Omkar, S., Mani, V., Karthikeyan, T., 2013. Multiobjective discrete particle swarm optimization for multisensor image alignment. IEEE Geosci. Remote Sens. Lett. 10 (5), 1095–1099.

Senthilnath, J., Prasad, R., 2014. A new sift matching criteria in a genetic algorithm framework for registering multisensory satellite imagery. In: Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing. ACM, p. 21.

Shechtman, E., Irani, M., 2007. Matching local self-similarities across images and videos. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on. IEEE, pp. 1–8.

Tissainayagam, P., Suter, D., 2005. Object tracking in image sequences using point features. Pattern Recogn. 38 (1), 105–113.

Tombari, F., Di Stefano, L., 2014. Interest points via maximal self-dissimilarities. In: Computer Vision–ACCV 2014. Springer, pp. 586–600.

Torabi, A., Najafianrazavi, M., Bilodeau, G.-A., 2011. A comparative evaluation of multimodal dense stereo correspondence measures. In: Robotic and Sensors Environments (ROSE), 2011 IEEE International Symposium on. IEEE, pp. 143–148.

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1. IEEE, pp. I–511.

Viola, P., Wells III, W.M., 1997. Alignment by maximization of mutual information. Int. J. Comput. Vis. 24 (2), 137–154.

Wachowiak, M.P., Smolíková, R., Zheng, Y., Zurada, J.M., Elmaghraby, A.S., 2004. An approach to multimodal biomedical image registration utilizing particle swarm optimization. IEEE Trans. Evol. Comput. 8 (3), 289–301.

Zhang, Z., Deriche, R., Faugeras, O., Luong, Q.-T., 1995. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Artif. Intell. 78 (1), 87–119.

Zitova, B., Flusser, J., 2003. Image registration methods: a survey. Image Vis. Comput. 21 (11), 977–1000.

Capítulo 5: Conclusiones y perspectivas futuras

5. Conclusiones y perspectivas futuras

En esta Tesis Doctoral se ha demostrado que las metodologías y procedimientos desarrollados contribuyen a la comunidad científica aportando una comparación clara entre las dos generaciones de dispositivos. Primero se realiza un análisis puramente metrológico, aportando unos resultados claros y concisos utilizables para ayudar a la decisión en la elección del dispositivo. Además, se añade una comparación en diferentes aplicaciones de visión artificial, haciendo énfasis en la reconstrucción 3D, la detección de objetos y la fusión de diferentes sensores, tres aplicaciones de claro interés, derivadas de los requisitos definidos durante la ejecución del proyecto SICEMAM y que se ven representadas en los artículos presentes en esta tesis. Con los análisis del rendimiento de estos dispositivos en diferentes aplicaciones de visión artificial se proporciona a la comunidad científica un conjunto de premisas sobre las que trabajar en futuras investigaciones. Durante el análisis de rendimiento se han ido creando diferentes herramientas, dispositivos y conjuntos de datos que aportaran facilidades e información a los futuros investigadores en la rama. En estas conclusiones se presentan las diferentes patentes y softwares que se han desarrollado y que se presentan en el APÉNDICE B: Patentes y el

APÉNDICE C: Software.

A continuación, tras alcanzar de forma satisfactoria los objetivos propuestos en la línea de investigación, se desarrollan en detalle las conclusiones derivadas de cada una de las publicaciones científicas y se describen las herramientas y dispositivos que se han generado. Estas conclusiones son complementadas con un desglose de las líneas de trabajo futuras.

5.1. Conclusiones

En relación con la **comparación metrológica** se obtiene una métrica y análisis de los resultados obtenidos por cada uno de los dispositivos. Durante la realización de este estudio metrológico se ha desarrollado una herramienta (GS-

MOD®) de captura con sensores de rango, en la que se puede incluir datos de calibración para obtener mejores resultados. Se comprueba de forma clara que la segunda generación obtiene mejores resultados que la primera. Se introduce una caracterización del dispositivo en función de la distancia, forma y ángulo de incidencia con precisión metrológica. Se llega a obtener una modelización del error provocado pudiendo ser este utilizado para prever el posible error en diferentes medidas.

En relación con la **comparación de rendimiento** en la tarea de visión artificial de **reconstrucción 3D**, se ha calculado una medida clara de la precisión en la reconstrucción de escenas u objetos, así como una medida real de la resolución. Se observa que a pesar de que la primera generación tiene una mayor resolución teórica en cuanto a número de pixeles, la segunda generación es capaz de obtener mejor resolución en capturas reales. Posiblemente esto se debe a la interpolación que hace en los cálculos la primera generación del dispositivo, siendo esto imposible de comprobar ya que este cálculo lo realiza un software privado no accesible. Todo el trabajo realizado aquí y el conocimiento obtenido se ha visto plasmado en la consecución de una patente con número P201531242: DISPOSITIVO AUTÓNOMO DE GENERACIÓN DE MODELOS FACIALES EN TRES DIMENSIONES.

En relación con la **comparación de rendimiento** en **reconocimiento 3D**, se ha observado que el dispositivo de segunda generación obtiene mejores resultados que el de primera generación, tanto en tasa de reconocimiento como en estimación del posicionamiento del objeto, tanto en ambientes despejados, como con oclusiones y para diferentes distancias y ángulos. Estos datos nos han ayudado a la elección de un algoritmo y dispositivo óptimo para la creación de una herramienta (Retrack®) para el reconocimiento de objetos con poca textura y en ambientes con oclusiones, que posteriormente se ha utilizado en el proyecto SICEMAM.

Respecto del **registro multimodal de imágenes**, se ha llegado a la conclusión de que el registro de imágenes provenientes de espectros electromagnéticos diferentes es un tema novedoso y poco estudiado para el que no existen demasiados datasets. Por lo tanto, se ha procedido a la generación de uno que cumpla los requerimientos para el análisis de las diferentes combinaciones

detector-descriptor para el registro de imágenes. Se ha podido avanzar en la combinación de detector y descriptor óptima según las modalidades que estemos registrando, permitiendo escoger esa combinación en la aplicación de cada caso. Debido a la dificultad que se ha encontrado durante este análisis, se concluye que el registro entre imágenes de distintas modalidades es extremadamente complejo debido a la no proporcionalidad de cambios de gradiente en las imágenes. Se ha observado que era necesaria la creación de un algoritmo capaz de descartar "outliers" e iterativamente seleccionar los puntos más adecuados hasta realizar el registro correcto entre las imágenes. Con la experiencia de registro entre imágenes obtenida durante este trabajo, se ha escrito otro artículo "*Infrared Cephalic-Vein to Assist Blood Extraction Tasks: Automatic Projection and Recognition*"[24], publicado en el congreso ISPRS INTERNATIONAL WORKSHOP "PHOTOGRAMMETRIC AND COMPUTER VISION TECHNIQUES FOR VIDEO SURVEILLANCE, BIOMETRICS AND BIOMEDICINE" – PSBB17, MAY 15-17, 2017, MOSCOW, RUSSIA. Además, ha proporcionado la idea para la creación de una patente P201331610, denominada "Plataforma fotogramétrica", que se basa en el registro de múltiples cámaras, perfectamente situadas para tener una cobertura de imagen visible de 360º.

5.2. Perspectivas futuras

Tras el desarrollo de esta Tesis Doctoral se han abierto diferentes líneas de investigación enfocadas a mejorar o complementar las metodologías propuestas.

Una línea de trabajo interesante que se abre a partir de la experiencia de la realización del primer trabajo, donde se identifica que una gran cantidad de tomas de datos nos permite hacer una mejor caracterización de los dispositivos, es el desarrollo de un prototipo que se ha patentado (BRazo Automatizado MUltisensor para la Reconstrucción 3D (BRAMUR3D)), pero que no ha sido posible su construcción por falta de medios. Este brazo nos puede permitir realizar adquisiciones de forma completamente automática, desde diferentes puntos de vista, con diferentes ángulos y distancias, ayudando a caracterizar de una forma más robusta los errores de las dos generaciones de sensores de rango.

Otra línea de investigación interesante sería evaluar la repetibilidad de medidas del sistema en función de la influencia de factores externos, como pueden ser la temperatura o reflectividad del material que se captura.

Con respeto al rendimiento en las diferentes aplicaciones, se podría estudiar la viabilidad en diferentes aplicaciones de visión artificial para seguir aumentando la caracterización de estos dispositivos, además de las ya introducidas en esta tesis, la reconstrucción 3D y el reconocimiento de objetos.

Otra rama por la que podría seguir investigando sería probar diferentes algoritmos o metodologías diferentes a las expuestas en esta Tesis Doctoral, como por ejemplo probar algoritmos de reconocimiento basados en características locales, por ejemplo, SHOT [25] el algoritmo que ha adquirido más fama dentro de esta categoría.

Finalmente, la idea que impulsó el desarrollo del tercer artículo, que se basa en el registro de imágenes multimodales, se podría extender para analizar la viabilidad de utilizar los métodos de registro multimodal estudiados para corregir los errores de radiometría en las nubes de puntos de los sensores de rango detectados en el segundo artículo de esta Tesis Doctoral.

Referencias

Referencias

1. Alcalde, J., *Marvin Minsky: "Las máquinas podrán hacer todo lo que hagan las personas, porque las personas sólo son máquinas"*, in *Muy Interesante*. 1996.
2. Aldoma, A., et al., *Point cloud library.* IEEE Robotics & Automation Magazine, 2012. **1070**(9932/12).
3. Newcombe, R.A., et al. *KinectFusion: Real-time dense surface mapping and tracking*. in *Mixed and augmented reality (ISMAR), 2011 10th IEEE international symposium on*. 2011. IEEE.
4. Karpathy, A., S. Miller, and L. Fei-Fei. *Object discovery in 3D scenes via shape analysis*. in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. 2013. IEEE.
5. Sturm, J., et al. *A benchmark for the evaluation of RGB-D SLAM systems*. in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. 2012. IEEE.
6. Khoshelham, K. and S.O. Elberink, *Accuracy and resolution of kinect depth data for indoor mapping applications.* Sensors, 2012. **12**(2): p. 1437-1454.
7. Gokturk, S.B., H. Yalcin, and C. Bamji. *A time-of-flight depth sensor-system description, issues and solutions*. in *Computer Vision and Pattern Recognition Workshop, 2004. CVPRW'04. Conference on*. 2004. IEEE.
8. Haggag, H., et al. *Measuring depth accuracy in RGBD cameras*. in *Signal Processing and Communication Systems (ICSPCS), 2013 7th International Conference on*. 2013. IEEE.
9. Lai, K., et al. *A large-scale hierarchical multi-view rgb-d object dataset*. in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. 2011. IEEE.
10. Gonzalez-Jorge, H., et al., *Metrological evaluation of microsoft kinect and asus xtion sensors.* Measurement, 2013. **46**(6): p. 1800-1806.
11. Aldoma, A., et al. *A global hypotheses verification method for 3d object recognition*. in *European conference on computer vision*. 2012. Springer.
12. Cheung, W. and G. Hamarneh. *N-sift: N-dimensional scale invariant feature transform for matching medical images*. in *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*. 2007. IEEE.
13. Hel-Or, Y., H. Hel-Or, and E. David, *Matching by tone mapping: Photometric invariant template matching.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2014. **36**(2): p. 317-330.
14. Wachinger, C. and N. Navab, *Entropy and Laplacian images: Structural representations for multi-modal registration.* Medical Image Analysis, 2012. **16**(1): p. 1-17.

15. Torabi, A., M. Najafianrazavi, and G.-A. Bilodeau. *A comparative evaluation of multimodal dense stereo correspondence measures*. in *Robotic and Sensors Environments (ROSE), 2011 IEEE International Symposium on*. 2011. IEEE.

16. Ashburner, J. and K. Friston, *Multimodal image coregistration and partitioning—a unified framework.* Neuroimage, 1997. **6**(3): p. 209-217.

17. Wachowiak, M.P., et al., *An approach to multimodal biomedical image registration utilizing particle swarm optimization.* Evolutionary Computation, IEEE Transactions on, 2004. **8**(3): p. 289-301.

18. Azuma, R.T., *A survey of augmented reality.* Presence: Teleoperators and virtual environments, 1997. **6**(4): p. 355-385.

19. Meta, Meta AR.

20. Microsoft. *Hololens*. [cited 2016; Available from: https://www.microsoft.com/microsoft-hololens/en-us.

21. Google. *Google Glass*. [cited 2016; Available from: https://www.google.com/glass/start/.

22. Tidop, G.; Available from: http://tidop.usal.es/es/.

23. Tidop, G. *Proyecto SICEMAM*. 20/07/2018]; Available from: http://tidop.usal.es/proyecto-siceman.

24. Lagüela, S., et al., *Infrared Cephalic-Vein to Assist Blood Extraction Tasks: Automatic Projection and Recognition.* The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017. **42**: p. 193.

25. Salti, S., F. Tombari, and L. Di Stefano, *SHOT: Unique signatures of histograms for surface and texture description.* Computer Vision and Image Understanding, 2014. **125**: p. 251-264.

APÉNDICE A: Indexación y factor de impacto de las publicaciones

Publicación 1: Metrological comparison between Kinect I and Kinect II sensors

MEASUREMENT

ISSN: 0263-2241

ELSEVIER SCI LTD
THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD OX5 1GB, OXON, ENGLAND
ENGLAND

Go to Journal Table of Contents Go to Ulrich's

Titles
ISO: Measurement
JCR Abbrev: MEASUREMENT

Categories
ENGINEERING, MULTIDISCIPLINARY - SCIE;
INSTRUMENTS & INSTRUMENTATION - SCIE;

Languages
ENGLISH

8 Issues/Year;

Key Indicators

Year	Total Cites Graph	Journal Impact Factor Graph	Impact Factor Without Journal Self Cites Graph	5 Year Impact Factor Graph	Immediacy Index Graph	Citable Items Graph	Cited Half-Life Graph	Citing Half-Life Graph	Eigenfacto Score Graph	Article Influence Score Graph	% Articles in Citable Items Graph	Normalizec Eigenfacto Graph	Average JIF Percentile Graph
2017	8,141	2.218	1.996	2.312	0.571	576	3.6	7.8	0.01…	0.446	99.13	1.89…	69.057
2016	6,513	2.359	2.018	2.255	0.464	784	3.3	8.2	0.01…	0.458	99.49	1.58…	76.618
2015	4,072	1.742	1.413	1.637	0.422	524	3.0	7.8	0.01…	0.386	97.33	1.18…	68.157
2014	3,037	1.484	1.168	1.424	0.415	547	3.3	8.3	0.00…	0.341	99.09	0.86…	66.371
2013	2,229	1.526	1.009	1.339	0.253	491	3.6	8.2	0.00…	0.284	96.33	0.55…	71.385
2012	1,575	1.130	0.807	1.159	0.296	294	3.8	8.2	0.00…	0.339	96.94	Not …	57.091
2011	1,110	0.836	0.702	0.854	0.202	252	4.6	8.8	0.00…	0.264	97.62	Not …	48.314
2010	1,017	0.853	0.661	0.885	0.154	195	5.2	9.4	0.00…	0.256	96.92	Not …	49.105
2009	869	0.761	0.602	0.933	0.164	195	5.6	>10.0	0.00…	0.294	97.95	Not …	46.295
2008	635	0.662	0.605	0.891	0.121	124	5.0	9.5	0.00…	0.306	99.19	Not …	36.600
2007	420	0.500	0.452	0.571	0.036	110	5.8	9.0	0.00…	0.202	100.00	Not …	39.417
2006	415	0.525	0.468	Not A…	0.087	103	5.7	9.4	Not A…	Not …	100.00	Not …	40.566
2005	289	0.413	0.349	Not A…	0.045	67	5.6	9.7	Not A…	Not …	100.00	Not …	38.750
2004	244	0.451	0.393	Not A…	0.068	74	5.7	9.3	Not A…	Not …	100.00	Not …	46.952
2003	228	0.434	0.377	Not A…	0.014	69	5.7	9.5	Not A…	Not …	98.55	Not …	35.714

Journal Impact Factor

Cites in 2015 to items published in: 2014 =868, 2013 =940, Sum: 1808
Number of items published in: 2014 =547, 2013 =491, Sum: 1038

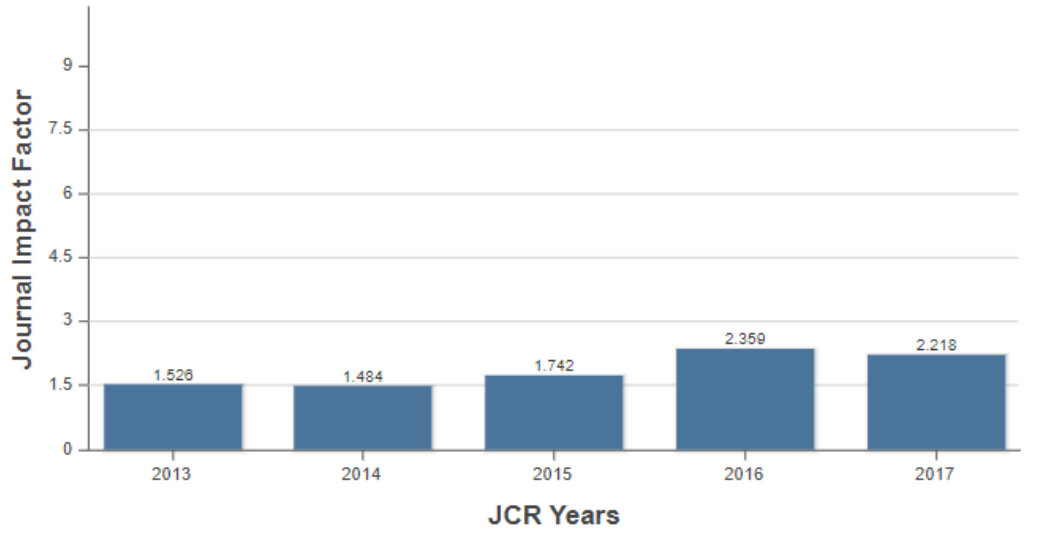
$$\text{Calculation} = \frac{\text{Cites to recent items}}{\text{Number of recent items}} = \frac{1808}{1038} = 1.742$$

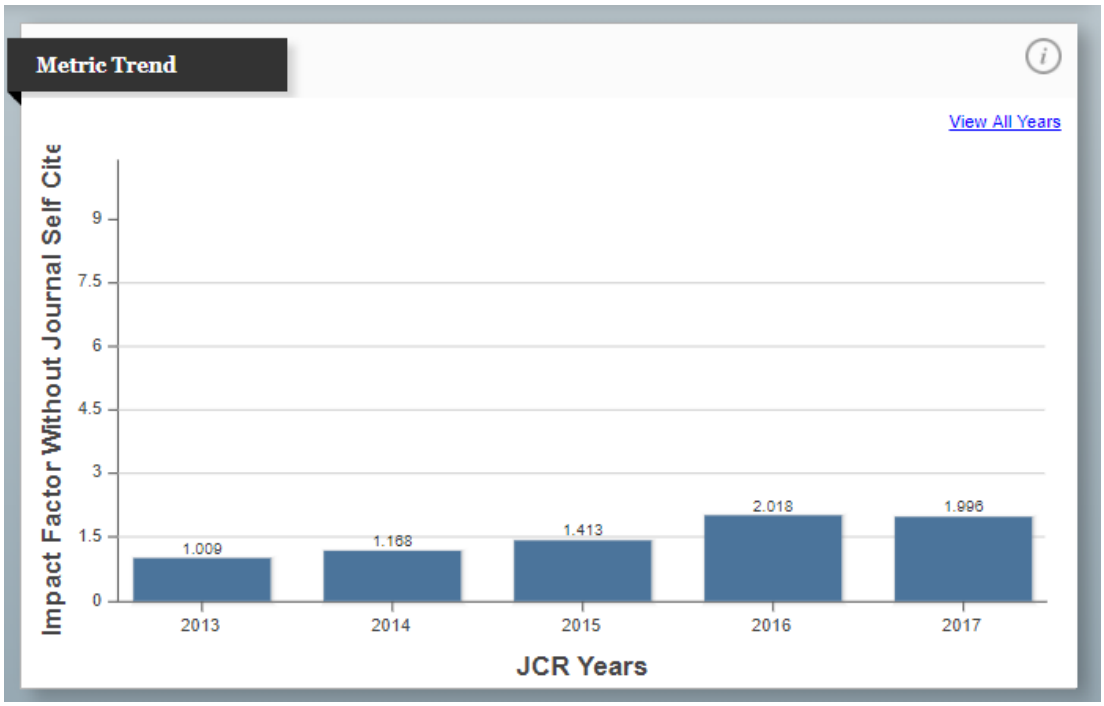
Journal Source Data

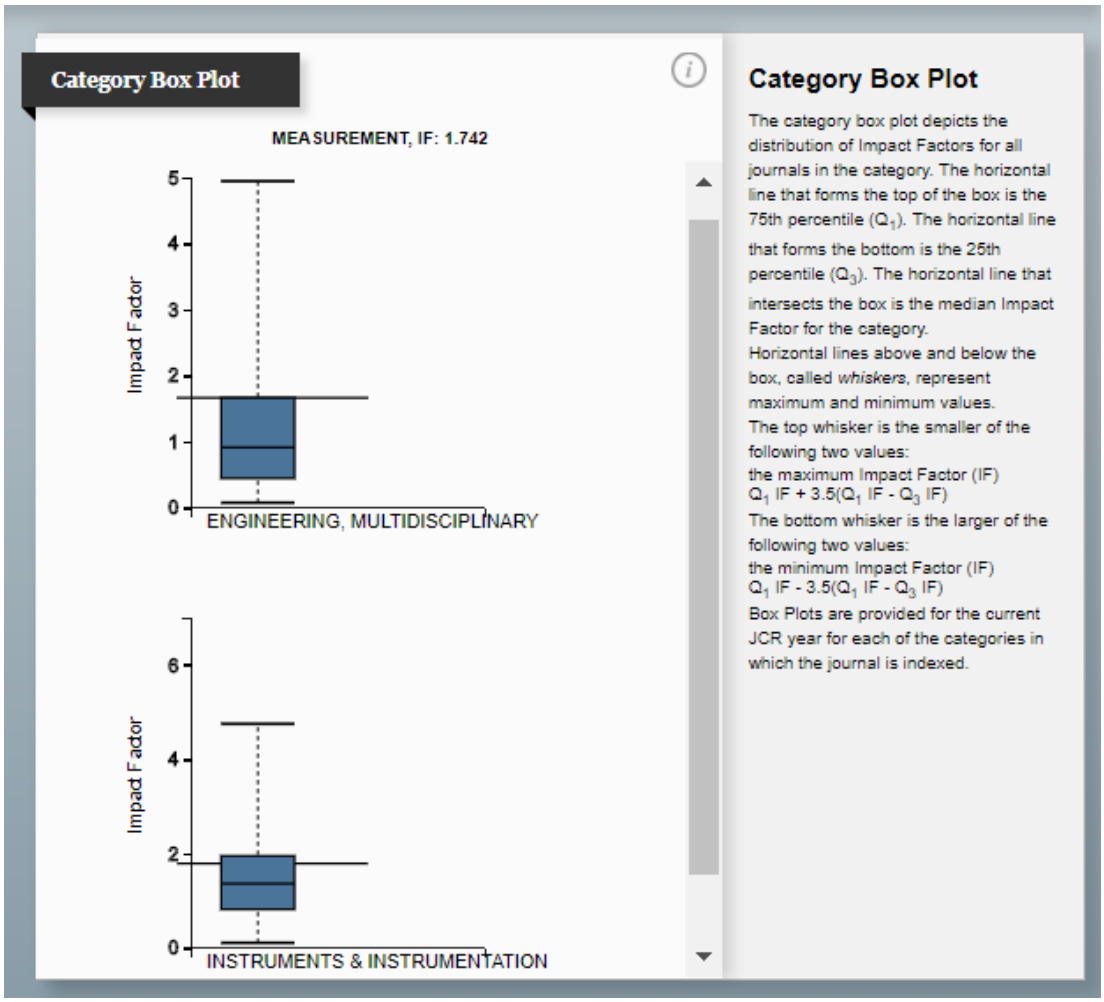
	Citable Items			Other (O)	Percentage (C/(C + O))
	Articles	Reviews	Combined (C)		
Number in JCR Year 2015...	510	14	524	2	99%
Number of References (B)	14,342	837	15,179	4	99%
Ratio (B/A)	28.1	59.8	29.0	2.0	

Metric Trend

View All Years




Metric Trend
View All Years
Impact Factor Without Journal Self Cite
9
7.5
6
4.5
3
1.5
0
1.009
1.168
1.413
2.018
1.996
2013
2014
2015
2016
2017
JCR Years


Category Box Plot
MEASUREMENT, IF: 1.742
Impact Factor
5
4
3
2
1
0
ENGINEERING, MULTIDISCIPLINARY
Impact Factor
6
4
2
0
INSTRUMENTS & INSTRUMENTATION
Category Box Plot
The category box plot depicts the distribution of Impact Factors for all journals in the category. The horizontal line that forms the top of the box is the 75th percentile (Q_1). The horizontal line that forms the bottom is the 25th percentile (Q_3). The horizontal line that intersects the box is the median Impact Factor for the category.
Horizontal lines above and below the box, called whiskers, represent maximum and minimum values.
The top whisker is the smaller of the following two values:
the maximum Impact Factor (IF)
Q_1 IF + 3.5(Q_1 IF - Q_3 IF)
The bottom whisker is the larger of the following two values:
the minimum Impact Factor (IF)
Q_1 IF - 3.5(Q_1 IF - Q_3 IF)
Box Plots are provided for the current JCR year for each of the categories in which the journal is indexed.

Publicación 2: Analysis and Evaluation Between the First and the Second Generation of RGB-D Sensors

IEEE SENSORS JOURNAL

ISSN: 1530-437X

IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC
445 HOES LANE,PISCATAWAY,USANJ 08855-4141
USA

Go to Journal Table of Contents Go to Ulrich's

Titles
ISO: IEEE Sens. J.
JCR Abbrev: IEEE SENS J

Categories
ENGINEERING, ELECTRICAL & ELECTRONIC - SCIE;
INSTRUMENTS & INSTRUMENTATION - SCIE;
PHYSICS, APPLIED - SCIE;

Languages
English

12 Issues/Year;

Key Indicators

Year	Total Cites Graph	Journal Impact Factor Graph	Impact Factor Without Journal Self Cites Graph	5 Year Impact Factor Graph	Immediacy Index Graph	Citable Items Graph	Cited Half-Life Graph	Citing Half-Life Graph	Eigenfacto Score Graph	Article Influence Score Graph	% Articles in Citable Items Graph	Normalizec Eigenfacto Graph	Average JIF Percentile Graph
2017	14,250	2.617	2.263	2.698	0.581	931	4.0	6.7	0.02...	0.564	98.39	3.19...	70.560
2016	11,167	2.512	2.052	2.527	0.456	1,059	4.0	7.0	0.02...	0.556	98.58	2.66...	71.942
2015	7,418	1.889	1.484	1.988	0.393	852	4.2	6.8	0.01...	0.549	98.36	2.20...	69.013
2014	6,154	1.762	1.522	1.901	0.326	534	4.1	6.8	0.01...	0.564	99.44	2.05...	65.106
2013	5,253	1.852	1.578	1.932	0.381	607	4.3	6.8	0.01...	0.537	99.34	1.70...	71.572
2012	3,961	1.475	1.285	1.758	0.348	469	4.5	7.3	0.01...	0.523	98.93	Not ...	60.701
2011	3,172	1.520	1.366	1.728	0.332	371	4.2	7.4	0.01...	0.523	99.73	Not ...	65.706
2010	2,590	1.473	1.380	1.590	0.265	264	4.0	7.2	0.01...	0.514	99.24	Not ...	64.190
2009	2,315	1.581	1.438	1.685	0.145	249	3.9	7.0	0.01...	0.566	100.00	Not ...	69.001
2008	1,968	1.610	1.478	1.818	0.174	282	3.6	6.6	0.01...	0.529	98.23	Not ...	66.746
2007	1,335	1.340	1.166	Not A...	0.210	233	3.3	6.8	0.00...	Not ...	99.57	Not ...	65.464
2006	834	1.117	0.976	Not A...	0.135	223	3.3	6.8	Not ...	Not ...	99.55	Not ...	61.063
2005	471	1.100	1.018	Not A...	0.065	186	2.9	7.6	Not ...	Not ...	100.00	Not ...	63.383

Journal Impact Factor

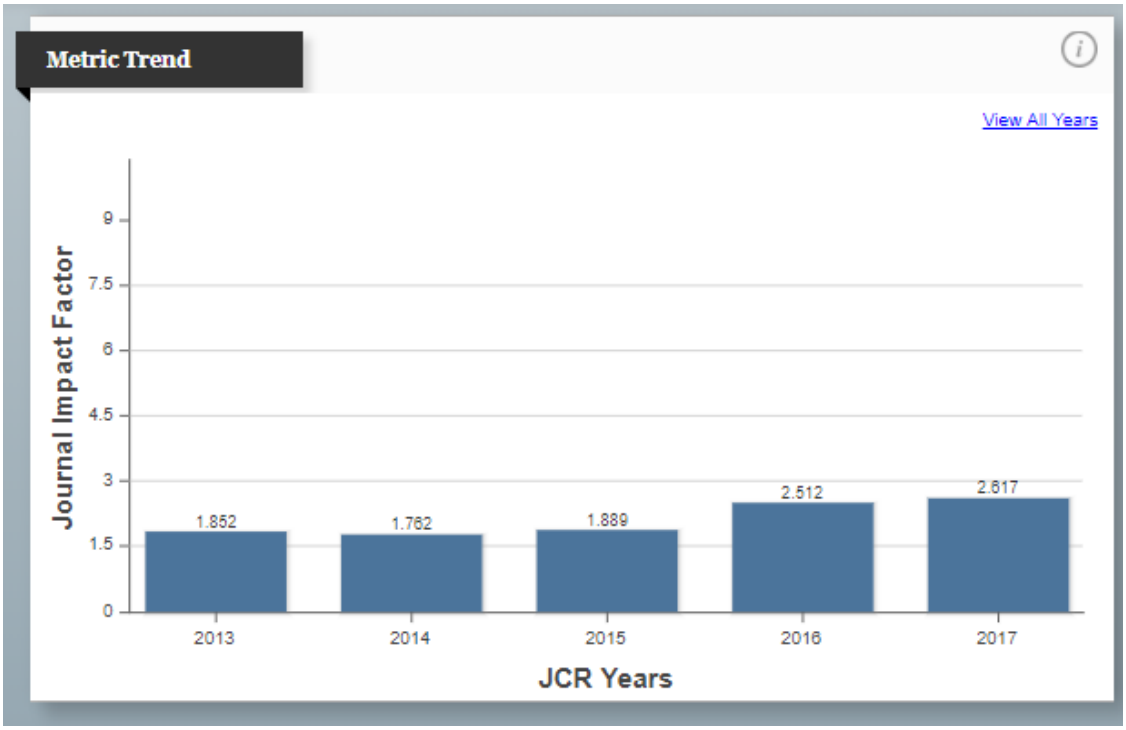
Cites in 2015 to items published in: 2014 =861 2013 =1298 Sum: 2159

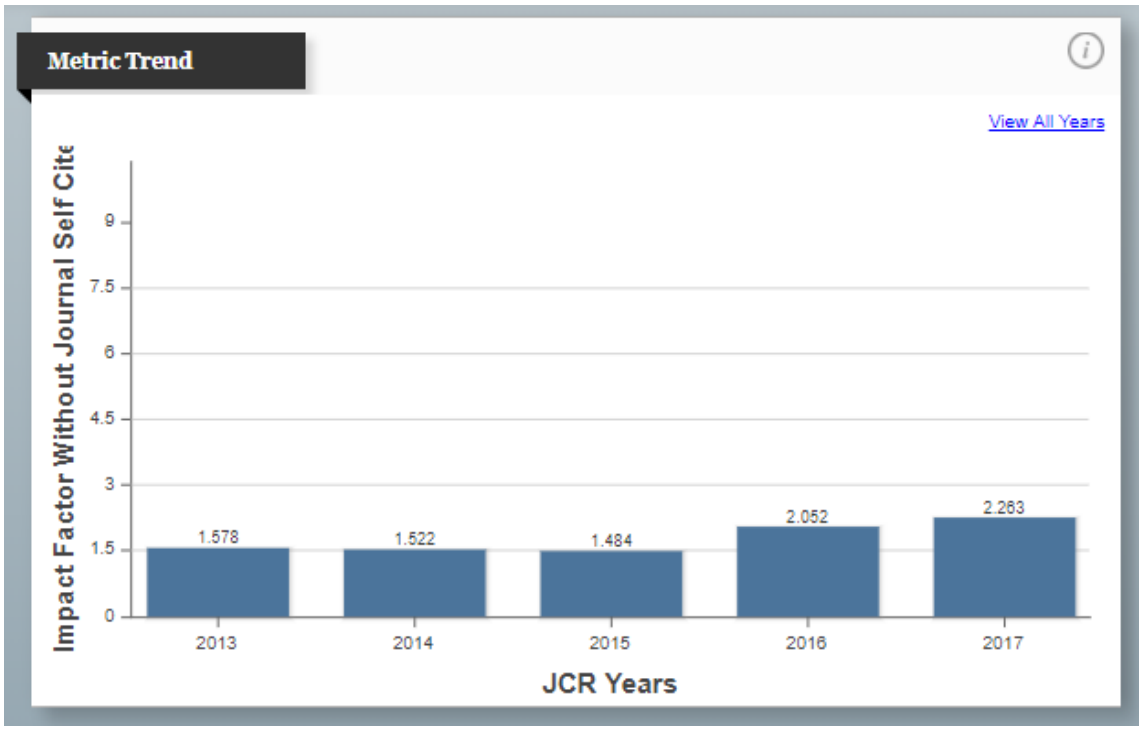
Number of items published in: 2014 =534 2013 =609 Sum: 1143

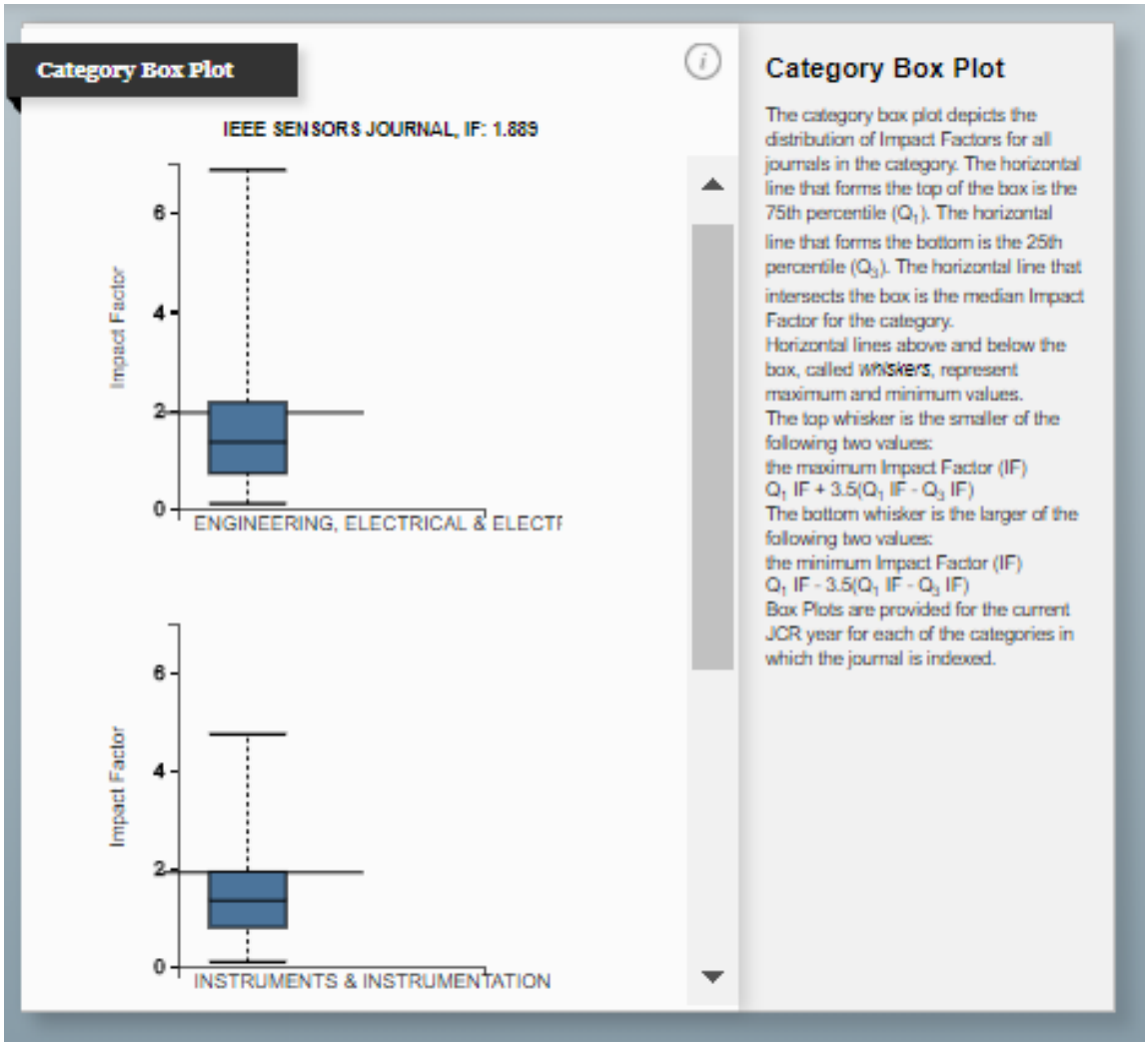
$$\text{Calculation} = \frac{\text{Cites to recent items}}{\text{Number of recent items}} = \frac{2159}{1143} = \mathbf{1.889}$$

Journal Source Data

	Citable Items			Other (O)	Percentage (C/(C + O))
	Articles	Reviews	Combined (C)		
Number in JCR Year 2015...	838	14	852	5	99%
Number of References (B)	23,381	1,158	24,539	0	100%
Ratio (B/A)	27.9	82.7	28.8	0.0	




Metric Trend
View All Years
Impact Factor Without Journal Self Cite
9
7.5
6
4.5
3
1.5
0
1.578
1.522
1.484
2.052
2.263
2013
2014
2015
2016
2017
JCR Years


Category Box Plot
IEEE SENSORS JOURNAL, IF: 1.889
Impact Factor
6
4
2
0
ENGINEERING, ELECTRICAL & ELECTR
Impact Factor
6
4
2
0
INSTRUMENTS & INSTRUMENTATION
Category Box Plot
The category box plot depicts the distribution of Impact Factors for all journals in the category. The horizontal line that forms the top of the box is the 75th percentile (Q_1). The horizontal line that forms the bottom is the 25th percentile (Q_3). The horizontal line that intersects the box is the median Impact Factor for the category.
Horizontal lines above and below the box, called whiskers, represent maximum and minimum values.
The top whisker is the smaller of the following two values:
the maximum Impact Factor (IF)
Q_1 IF + 3.5(Q_1 IF - Q_3 IF)
The bottom whisker is the larger of the following two values:
the minimum Impact Factor (IF)
Q_1 IF - 3.5(Q_1 IF - Q_3 IF)
Box Plots are provided for the current JCR year for each of the categories in which the journal is indexed.

Publicación 3: Analysis and Evaluation Between the First and the Second Generation of RGB-D Sensors

ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING

ISSN: 0924-2716

ELSEVIER SCIENCE BV
PO BOX 211,1000 AE AMSTERDAM,NETHERLANDS
NETHERLANDS

Go to Journal Table of Contents Go to Ulrich's

Titles
ISO: ISPRS-J. Photogramm. Remote Sens.
JCR Abbrev: ISPRS J PHOTOGRAMM

Categories
GEOGRAPHY, PHYSICAL - SCIE; GEOSCIENCES, MULTIDISCIPLINARY - SCIE; REMOTE SENSING - SCIE; IMAGING SCIENCE & PHOTOGRAPHIC TECHNOLOGY - SCIE;

Languages
Multi-Language

12 Issues/Year;

Key Indicators

Year	Total Cites Graph	Journal Impact Factor Graph	Impact Factor Without Journal Self Cites Graph	5 Year Impact Factor Graph	Immediacy Index Graph	Citable Items Graph	Cited Half-Life Graph	Citing Half-Life Graph	Eigenfactor Score Graph	Article Influence Score Graph	% Articles in Citable Items Graph	Normalized Eigenfactor Graph	Average JIF Percentile Graph
2017	8,535	5.994	5.249	6.592	0.944	197	4.7	7.2	0.01...	1.509	98.98	1.84...	95.585
2016	7,720	6.387	5.624	6.457	1.108	167	4.8	7.4	0.01...	1.362	89.82	1.54...	98.102
2015	5,125	4.188	3.608	5.062	0.978	182	5.0	7.7	0.01...	1.270	98.35	1.22...	93.478
2014	4,120	3.132	2.623	4.652	0.890	172	5.5	7.4	0.00...	1.062	96.51	0.86...	83.467
2013	3,088	2.902	2.398	4.202	0.366	142	5.9	7.9	0.00...	1.171	96.48	0.77...	83.841
2012	2,496	3.313	2.843	4.026	0.470	100	5.9	7.9	0.00...	1.209	98.00	Not ...	88.522
2011	1,879	2.885	2.508	3.435	0.323	93	6.2	7.5	0.00...	1.028	98.92	Not ...	85.659
2010	1,532	2.184	1.947	2.908	0.481	54	6.6	6.2	0.00...	0.959	98.15	Not ...	78.728
2009	1,328	2.308	2.076	3.267	0.412	68	6.3	7.9	0.00...	0.953	95.59	Not ...	80.861
2008	1,235	2.293	2.032	3.694	0.283	46	6.3	7.4	0.00...	0.886	97.83	Not ...	77.716
2007	769	1.116	0.971	2.356	0.133	45	6.8	7.7	0.00...	0.937	97.78	Not ...	53.237
2006	632	1.381	1.238	Not ...	0.106	47	7.2	5.8	Not ...	Not ...	100.00	Not ...	66.265
2005	447	1.674	1.604	Not ...	0	22	6.4	6.4	Not ...	Not ...	95.45	Not ...	78.241
2004	452	1.317	1.250	Not ...	0.200	20	5.6	7.6	Not ...	Not ...	100.00	Not ...	66.500
2003	301	0.472	0.452	Not ...	0.130	23	5.2	5.8	Not ...	Not ...	100.00	Not ...	24.062
2002	231	0.389	0.361	Not ...	0.081	37	4.4	5.5	Not ...	Not ...	100.00	Not ...	22.480
2001	169	0.963	0.759	Not ...	0.062	16	4.3	5.7	Not ...	Not ...	100.00	Not ...	64.913

Journal Impact Factor

Cites in 2017 to items published in: 2016 =811, 2015 =1281, Sum: 2092

Number of items published in: 2016 =167, 2015 =182, Sum: 349

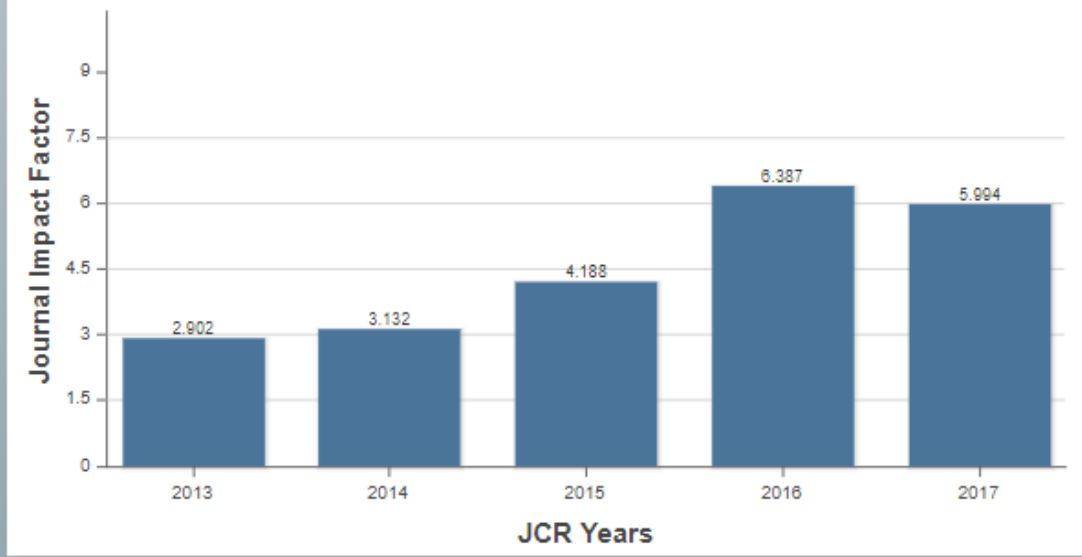
$$\text{Calculation} = \frac{\text{Cites to recent items}}{\text{Number of recent items}} \quad \frac{2092}{349} = 5.994$$

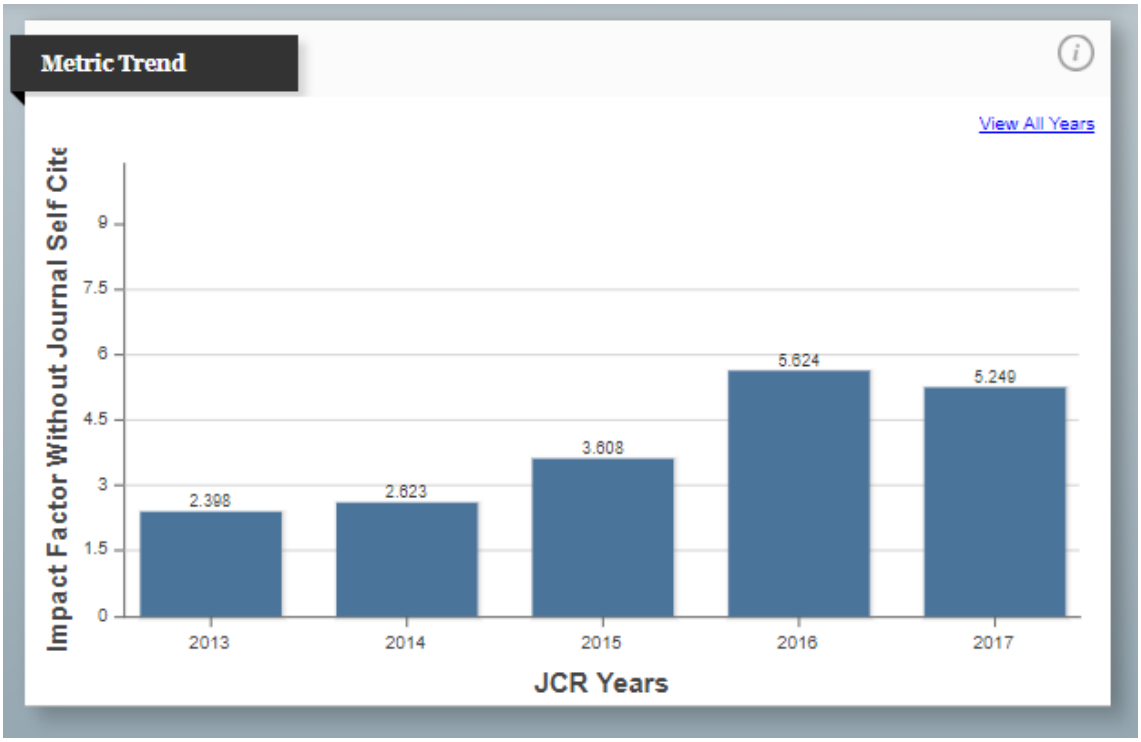
Journal Source Data

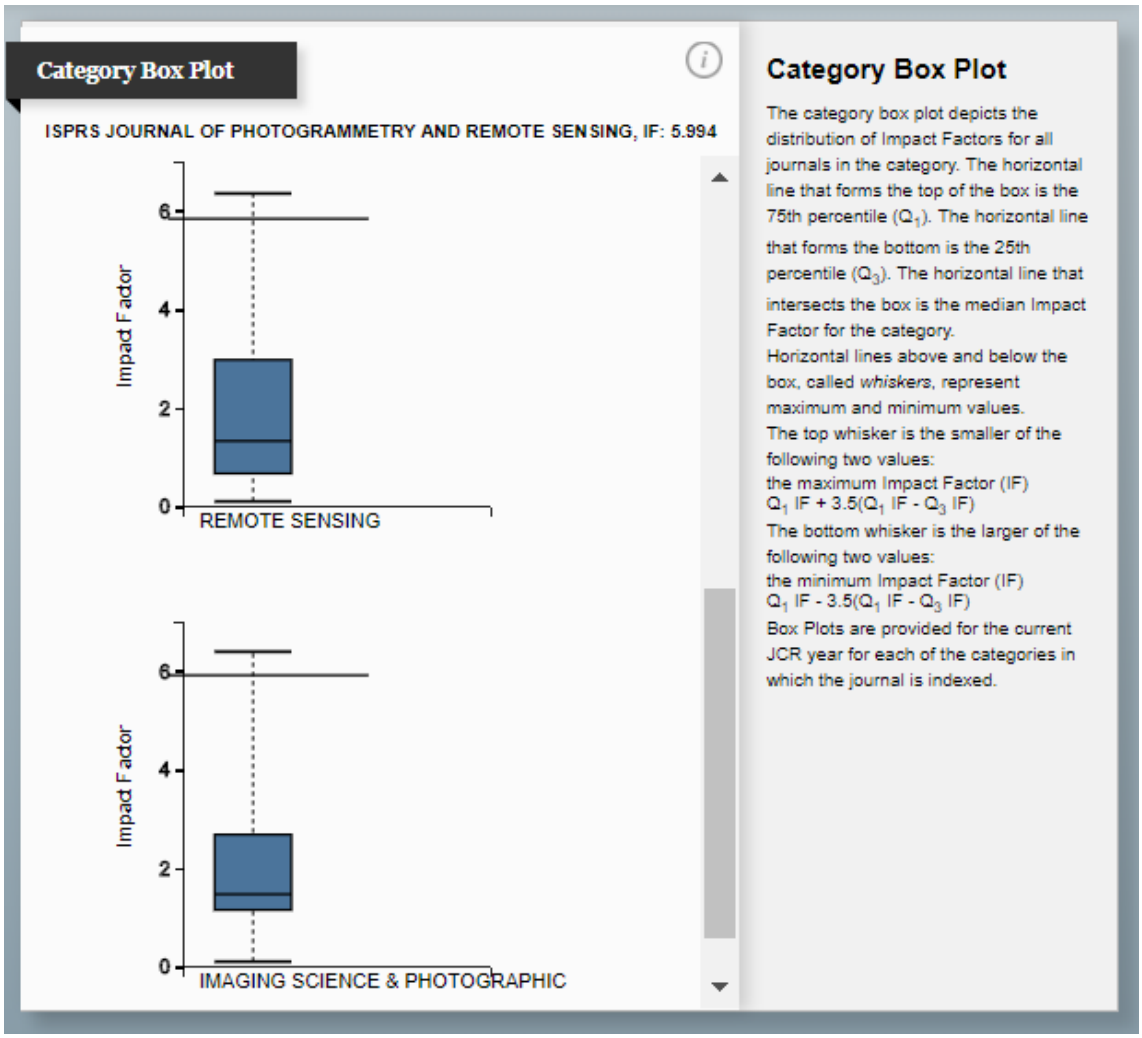
	Citable Items			Other (O)	Percentage (C/(C + O))
	Articles	Reviews	Combined (C)		
Number in JCR Year 2017...	195	2	197	1	99%
Number of References (B)	10,620	281	10,901	9	99%
Ratio (B/A)	54.5	140.5	55.3	9.0	

Metric Trend

View All Years




Metric Trend
View All Years
Impact Factor Without Journal Self Cite
9
7.5
6
4.5
3
1.5
0
2.398
2.623
3.608
5.624
5.249
2013
2014
2015
2016
2017
JCR Years


Category Box Plot
ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING, IF: 5.994
Impact Factor
6
4
2
0
REMOTE SENSING
Impact Factor
6
4
2
0
IMAGING SCIENCE & PHOTOGRAPHIC
Category Box Plot
The category box plot depicts the distribution of Impact Factors for all journals in the category. The horizontal line that forms the top of the box is the 75th percentile (Q_1). The horizontal line that forms the bottom is the 25th percentile (Q_3). The horizontal line that intersects the box is the median Impact Factor for the category.
Horizontal lines above and below the box, called whiskers, represent maximum and minimum values.
The top whisker is the smaller of the following two values:
the maximum Impact Factor (IF)
Q_1 IF + 3.5(Q_1 IF - Q_3 IF)
The bottom whisker is the larger of the following two values:
the minimum Impact Factor (IF)
Q_1 IF - 3.5(Q_1 IF - Q_3 IF)
Box Plots are provided for the current JCR year for each of the categories in which the journal is indexed.

APÉNDICE B: Patentes

Patente 1: Plataforma fotogramétrica

MINISTERIO DE INDUSTRIA, ENERGIA Y TURISMO

Oficina Española de Patentes y Marcas

Nº SOLICITUD: **P201331610**
Nº PUBLICACIÓN: **ES2535205**
TITULAR/ES:
UNIVERSIDAD DE SALAMANCA
UNIVERSIDAD DE CASTILLA LA MANCHA

FECHA EXPEDICIÓN: 17/02/2016

TÍTULO
DE
PATENTE DE INVENCIÓN

Cumplidos los requisitos previstos en la vigente Ley 11/1986, de 20 de marzo, de Patentes, se expide el presente TÍTULO, acreditativo de la concesión de la Patente de Invención. La solicitud ha sido tramitada y concedida con realización del Informe sobre el Estado de la Técnica y **sin examen previo** de los requisitos sustantivos de patentabilidad.

Se otorga al titular un derecho de exclusiva en todo el territorio nacional, bajo las condiciones y con las limitaciones previstas en la Ley de Patentes. La duración de la patente será de **veinte años** contados a partir de la fecha de presentación de la solicitud (04/11/2013).

La patente se concede sin perjuicio de tercero y sin garantía del Estado en cuanto a la validez y a la utilidad del objeto sobre el que recae.

Para mantener en vigor la patente concedida, deberán abonarse las tasas anuales establecidas, que se pagarán por años adelantados. Asimismo, deberá explotarse el objeto de la invención, bien por su titular o por medio de persona autorizada de acuerdo con el sistema de licencias previsto legalmente, dentro del plazo de cuatro años a partir de la fecha presentación de la solicitud de patente, o de tres años desde la publicación de la concesión en el Boletín Oficial de la Propiedad Industrial, aplicándose el plazo que expire más tarde.

OFICINA ESPAÑOLA DE PATENTES Y MARCAS

Fdo.: Ana María Redondo Mínguez
Jefe/a de Servicio de Actuaciones Administrativas
(P.D. del Director del Departamento de Patentes e I.T., resolución 05/09/2007)

(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 535 205**

(21) Número de solicitud: 201331610

(51) Int. CI.:

G03B 17/56 (2006.01)
G03B 35/00 (2006.01)

(12) PATENTE DE INVENCIÓN B1

(22) Fecha de presentación:
04.11.2013

(43) Fecha de publicación de la solicitud:
06.05.2015

Fecha de la concesión:
17.02.2016

(45) Fecha de publicación de la concesión:
24.02.2016

(73) Titular/es:
UNIVERSIDAD DE SALAMANCA (85.0%)
Patio de Escuelas, 1
37008 Salamanca (Salamanca) ES y
UNIVERSIDAD DE CASTILLA LA MANCHA (15.0%)

(72) Inventor/es:
RODRÍGUEZ GONZÁLVEZ, Pablo;
GESTO DÍAZ, Manuel;
FERNÁNDEZ HERNÁNDEZ , Jesús;
GONZÁLEZ AGUILERA, Diego;
HOLGADO BARCO, Alberto y
HERNÁNDEZ LÓPEZ, David

(74) Agente/Representante:
PONS ARIÑO, Ángel

(54) Título: **Plataforma fotogramétrica**

(57) Resumen:
Plataforma fotogramétrica que permite la instalación de una pluralidad de cámaras (4) en una estructura (1) de soporte con al menos unas paredes laterales (10) y una base (12). Comprende unos alojamientos (2) en la estructura (1) de soporte en los que se encuentran unas bahías (3) destinadas a recibir las cámaras (4). También comprende al menos un soporte adicional (5) unido de forma abatible a la estructura (1) y que también dispone de una bahía (3). Son también parte de la plataforma un sistema de sujeción (6) destinado a fijar la inclinación del soporte adicional (5) respecto a la estructura (1) y un elemento de conexión (7) dispuesto en la base (12) destinado a recibir un tubo de conexión (8) destinado a permitir su colocación en un dispositivo externo que puede ser aéreo o terrestre.

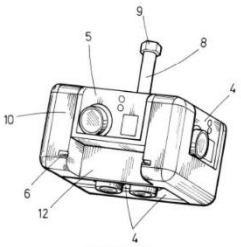


FIG.5

ES 2 535 205 B1

ES 2 535 205 B1

PLATAFORMA FOTOGRAMÉTRICA

D E S C R I P C I Ó N

OBJETO DE LA INVENCIÓN

La presente invención se puede incluir en el campo técnico de las plataformas fotogramétricas. Concretamente ésta es multifunción y permite realizar tanto tomas aéreas como terrestres.

ANTECEDENTES DE LA INVENCIÓN

La fotogrametría es una técnica que se emplea para determinar las propiedades geométricas de los objetos y las situaciones espaciales a partir de imágenes fotográficas.

En la actualidad se ha generalizado el empleo de vehículos aéreos no tripulados de bajo coste (UAV-*unmanned aerial vehicles*) a partir de lo cual ha proliferado el diseño de plataformas aéreas para sustentar cámaras y diversos sensores con diferentes propósitos fotogramétricos.

De forma coincidente en el tiempo, la miniaturización de sensores de posicionamiento y navegación, junto con la aparición de cámaras fotográficas de bajo coste y peso, ha permitido su embarcación no sólo en plataformas aéreas UAV sino también en plataformas terrestres estáticas o móviles. Esto permite configuraciones multi-sensor y multi-cámara capaces de garantizar resultados de gran calidad e incluso superar en calidad y resolución a las imágenes capturadas por los medios aéreos tripulados y los modernos sistemas de escaneo láser terrestre.

En el campo de los UAV, las plataformas para sustentar cámaras existentes hasta la fecha están especializadas, desde el punto de vista fotogramétrico, en dos líneas de acción hasta ahora incompatibles:

(i) la captura de imágenes panorámicas con fines de inspección y fotointerpretación

(mediante el disparo simultáneo de varias cámaras sincronizadas);

(ii) la adquisición de imágenes estereoscópicas verticales y oblicuas para la reconstrucción con fines métricos de cualquier escenario u objeto (mediante el empleo de dos cámaras de disparo síncrono en disposición estereoscópica de base conocida, o una sola disparando en diferentes posiciones según el avance del propio vehículo aéreo).

En esta segunda línea de acción, la adquisición fotográfica con una sola cámara está limitada a la separación (baselínea) entre las cámaras y por ende a sus posibles fuentes de error. Actualmente, para minimizar ese problema se emplea una base conocida, pequeña y calibrada, que permite un mayor grado de solape y la posibilidad de realizar pares múltiples (*multiple dense stereo matching*). Esto redunda en una mejora de la precisión de las reconstrucciones.

Del estado de la técnica se conocen plataformas fotogramétricas que tienen ciertas limitaciones. Por ejemplo se conocen plataformas fotogramétricas que solo se pueden emplear en vehículos aéreos y que además en la disposición en la zona nadiral de la plataforma ésta recibe muchas distorsiones debidas a la geometría de "ojo de pez" de las lentes (y su distribución). Otra desventaja es que no permite estereoscopía de base conocida.

Otras plataformas conocidas se instalan mediante sistemas de contrapeso y no permiten la captura de panoramas de forma sincronizada sino que es necesario rotar el propio vehículo aéreo en el que se instalan.

También se conocen plataformas que están restringidas a un uso solo terrestre, o tienen problemas con el panorama que resulta ya que no proporciona cobertura cenital completa. En otras no se permite la captura esteroscópica síncrona de base conocida o se permite solo en dirección horizontal.

DESCRIPCIÓN DE LA INVENCIÓN

La presente invención ofrece una plataforma fotogramétrica que permite realizar una

captura simultánea de imágenes panorámicas y estereoscópicas. La plataforma está diseñada para la colocación estratégica de una pluralidad de cámaras para la obtención de panoramas y prestando especial atención a la toma estereoscópica simple y múltiple que permita la automatización en la reconstrucción 3D fotogramétrica de las imágenes obtenidas.

Para ello, la plataforma presenta una configuración optimizada que minimiza la excentricidad entre las cámaras que se colocan en ella y distribuye adecuadamente el centro de masas para una mayor estabilidad en el caso de ser instalada en un vehículo aéreo no tripulado.

Por otro lado, la presente invención no está limitada a su aplicabilidad en plataformas aéreas no tripuladas (UAV), ya que no siempre es posible su utilización aérea (restricciones de seguridad o físicas); pudiendo ser embarcada en cualquier soporte o vehículo terrestre tanto en modo estático (embarcada en jalón o trípode) como dinámico (embarcada en sistemas terrestres móviles).

La plataforma comprende una estructura en la que hay una pluralidad de paredes laterales dispuestas con un ángulo de 90º entre sí y que son ortogonales a una base. Comprende también una pieza de soporte adicional, que está unida a la base y que es abatible. Dicho soporte adicional puede pasar de una primera posición en la que está alineado con la base a una segunda posición en la que queda ortogonal a la base, dispuesto entre dos de las paredes laterales.

En cada pared lateral y en el soporte adicional se encuentra una cámara y en la base se encuentran dos cámaras situadas contrapuestas entre sí. Además comprende un elemento de conexión con un tubo de conexión destinado a permitir su fijación a un dispositivo externo en el que se instala.

Así pues la plataforma permite su fácil intercambiabilidad entre una configuración aérea y una configuración terrestre para conseguir panoramas mediante la cobertura solapada de las cámaras dispuestas en ella (cámaras de disparo sincronizado). Adicionalmente se pueden obtener imágenes estereoscópicas en techos y bóvedas.

Además, en la configuración terrestre, se puede disponer el soporte adicional en la primera posición para realizar tomas múltiples estéreo y estéreo-panoramas, de gran interés estos últimos en las visitas virtuales inmersivas 3D. Para trabajar en esta configuración hay que girar la plataforma de forma acimutal hasta encuadrar el objeto de interés (alzados, fachadas) mediante el abatimiento del soporte adicional y la rotación sobre su propio eje vertical.

La posibilidad de realizar tomas fotográficas simultáneas con fines de fotointerpretación y fotogrametría y poder hacerlo tanto desde tierra como desde el aire (además de poder embarcar diferentes tipos de cámaras (compacta visible, infrarroja, termográfica, etc.)) permite abrir el camino a numerosas aplicaciones de la presente invención.

Más concretamente, la plataforma fotogramétrica permite obtener imágenes panorámicas para utilizarlas con fines de interpretación y análisis visual, así como imágenes estereoscópicas que se requieran en aquellas aplicaciones con fines métricos en forma de modelos 3D y ortofotos.

El gran solape existente entre las fotos resultantes (con posibilidad de configuración de diferentes baselíneas, corta y larga) unido a la redundancia procedente de los múltiples pares estéreo (imágenes estereoscópicas), permite garantizar una gran calidad en los procesos de matching (que se realiza comparando punto por punto) y por consiguiente en la obtención de las nubes de puntos 3D para la obtención de la imagen 3D final.

Con esta invención se aporta un avance en la adquisición de información fotogramétrica, con independencia del soporte sobre el que se monte la plataforma (soporte aéreo y/o terrestre) y el método fotogramétrico empleado (panorámico o estereoscópico).

Las principales áreas de aplicación de la plataforma fotogramétrica descrita son el análisis dimensional y reconstrucción tridimensional de edificios/fachadas tanto históricos

como de nueva construcción, el análisis dimensional y reconstrucción tridimensional de estructuras en ingeniería civil, la documentación y el cartografiado de yacimientos arqueológicos, y la catalogación e inventariado en urbanismo.

Así pues, en base a lo anteriormente descrito, las mayores ventajas de la presente invención son:

-Se puede emplear tanto en vehículos aéreos no tripulados como terrestres, adaptándose a sus características (como por ejemplo la disposición del centro de gravedad para embarcarse con estabilidad en plataformas aéreas no tripuladas). Además cuando se emplea en aplicaciones terrestres se puede colocar tanto en dispositivos móviles (vehículos rodados) como en dispositivos estáticos (del tipo trípodes, jalones, mochilas de transporte, etc.).

-Permite la captura simultánea de panoramas e imágenes múltiples estereoscópicas.

-Tiene un bajo coste.

-Es ligero, lo cual es un aspecto crucial en su vertiente aérea no tripulada con un peso del soporte estimado en 870gr con las 6 cámaras embarcadas.

-Es portable y robusto dada su ligereza y tiene un diseño compacto.

DESCRIPCIÓN DE LOS DIBUJOS

Para complementar la descripción que se está realizando y con objeto de ayudar a una mejor comprensión de las características de la invención, de acuerdo con un ejemplo preferente de realización práctica de la misma, se acompaña como parte integrante de dicha descripción, un juego de dibujos en donde con carácter ilustrativo y no limitativo, se ha representado lo siguiente:

Figura 1.- Muestra una vista en perspectiva de la estructura de soporte con las bahías en las que se alojan las cámaras.

Figura 2.- Muestra una vista en perspectiva de la plataforma fotogramétrica en la que el tubo de conexión está dispuesto perpendicular a la base y el soporte adicional está en la primera posición.

Figura 3.- Muestra una vista en perspectiva de la plataforma fotogramétrica en la que el tubo de conexión está dispuesto paralelo a la base y el soporte adicional está en la primera posición.

Figura 4.- Muestra una vista desde una perspectiva diferente de la plataforma fotogramética mostrada en la figura 3 en la que se aprecian las cámaras que se encuentran alojadas en la base.

Figura 5.- Muestra una vista en perspectiva de la plataforma fotogramétrica en la que el tubo de conexión está dispuesto perpendicular a la base y el soporte adicional está en la segunda posición.

REALIZACIÓN PREFERENTE DE LA INVENCIÓN

A continuación se describe con ayuda de las figuras 1 a 5 una realización preferente de la invención.

Se propone una plataforma fotogramétrica del tipo de las que permiten la instalación de una pluralidad de cámaras (4) en una estructura (1) de soporte, como la mostrada en la figura 1, con al menos unas paredes laterales (10) y una base (12).

Asimismo, la plataforma fotogramétrica comprende unos alojamientos en la estructura (1) de soporte en los que se encuentran unas bahías (3) destinadas a recibir las cámaras (4) y al menos un soporte adicional (5) unido de forma abatible a la estructura (1) y que también comprende una bahía (3). Además comprende un sistema de sujeción (6) destinado a fijar la inclinación de dicho soporte adicional (5) respecto a la estructura (1). Dicho sistema comprende unos pasadores que fijan la posición del soporte adicional (5) respecto a la estructura (1).

En función de cómo se coloque la cámara (4) en la bahía (3) del soporte adicional (5) se pueden obtener imágenes con pares estéreo de mayor o menor baselínea.

La plataforma fotogramétrica está destinada a ser instalada en un dispositivo externo. Una de las ventajas de la presente invención es que se puede instalar tanto en vehículos aéreos (tripulados y no tripulados) como en dispositivos terrestres. Para ello comprende un elemento de conexión (7) dispuesto en la estructura (1) de soporte destinado a recibir un tubo de conexión (8) destinado a permitir su colocación en el dispositivo externo en el que se instala.

El elemento de conexión (7) está dispuesto en el centro de masas de la estructura (1) y esto permite mejorar la estabilidad respecto al vehículo en las tomas aéreas.

Preferentemente la estructura (1) de soporte está fabricada en plástico duro (PVC) para optimizar la relación entre el peso (crítico en los vehículos aéreos no tripulados-UAV) y la resistencia.

En una realización de la invención, en la estructura (1) las caras laterales (10) y la base (12) están dispuestas ortogonalmente entre sí creando un espacio libre (11) entre ellas en el que está dispuesto el elemento de conexión (7).

En cada cara lateral (10) se encuentra una bahía (3) y en la base (12) se encuentran dos bahías (3) contrapuestas. Es decir, están dispuestas en el mismo plano con una orientación de 90º entre sí. Esta distribución permite la generación de pares estereoscópicos. En función de cómo se coloquen las cámaras (4) en estas bahías contrapuestas (3) se podrá contar con pares estéreo de mayor o menor baselínea.

Como se aprecia en la figura 1, preferentemente hay seis bahías (3) (una en cada pared lateral (10), dos en la base (12) y otra en el soporte adicional (5)).

Las bahías (3) pueden comprender unas guías que facilitan la colocación de las cámaras (4). Dichas cámaras (4) pueden ser cámaras deportivas (estilo GoPro®), así como otros

tipos de cámaras compactas.

Las bahías de la base (12) están preferentemente contrapuestas entre sí para garantizar el solape incluso para cámaras cuyo campo de vista vertical sea ligeramente menor de 90º y establecer simultáneamente una baselínea estereoscópica.

El soporte adicional (5) está unido a la base (12) dispuesto entre dos paredes laterales (10) y se desplaza entre una primera posición en la que está alineado con dicha base (12) y una segunda posición en la que queda ortogonal a la base (12) enfrentado a una tercera pared lateral (10).

Según esta realización, cuando el soporte adicional (5) está en la segunda posición, cuatro de las bahías (3) (las tres bahías de las paredes laterales (10) y la del soporte adicional (5)) se disponen según una configuración específica conformando un "anillo". Es decir, el perímetro de la estructura (1) queda cubierto con una bahía (3) en cada una de las paredes laterales (10) que están separadas 90º entre sí y con la bahía (3) del soporte adicional (5).

Las dos bahías (3) de la base (12) están en un plano ortogonal a las bahías de las paredes laterales (10). La distribución de las bahías (3) tiene un diseño óptimo para minimizar la excentricidad entre las cámaras (4) (paralaje) y garantizar una captura de imágenes individuales con solape suficiente para la realización de panoramas.

La plataforma fotogramétrica descrita consigue su dualidad aire-tierra mediante el soporte adicional (5) gracias a lo que la bahía (3) que está dispuesta en él tiene posibilidad de abatimiento. Así pues está permitida una rotación de 90º para cambiar su plano de disposición (de vertical a horizontal), pasando de la primera posición a la segunda posición descrita y viceversa.

Mediante la actuación del soporte adicional (5) para que pase a la primera posición, se dispone de tres bahías en las paredes laterales (10) destinadas a alojar unas cámaras (4) con las que componer hasta tres pares estereoscópicos separados por

sus respectivas baselíneas.

El elemento de conexión (7) es preferentemente una rosca doble ortogonal que dispone de un primer orificio (7.1) orientado hacia el soporte adicional (5) y un segundo orificio (7.2) orientado ortogonal a éste paralelo a la base (12). El tubo de conexión (8) se puede colocar en cualquiera de los orificios (7.1, 7.2) para poder instalarlo en el dispositivo externo en el que se quiere emplear (UAV, trípode, vehículo terrestre...). A este respecto el tubo de conexión (8) está preparado para recibir unas roscas de empalme (9) necesarias en cada caso en función del dispositivo externo en el que se quiera instalar.

En las figuras 2 y 5 se observan realizaciones en las que el tubo de conexión (8) está dispuesto en el segundo orificio (7.2). En las figuras 3 y 4 se observa una realización en la que el tubo de conexión (8) está dispuesto en el primer orificio (7.1).

Asimismo la plataforma fotogramétrica puede llevar al menos una tapa (2) que se une a la estructura (1) de soporte en las bahías (3), destinada a asegurar la posición de las cámaras (4) en las bahías (3). Preferentemente dicha tapa (2) tiene una configuración en "U" y se dispone sobre los extremos de las paredes laterales (10) garantizando la integridad de la estructura (1) de soporte y permitiendo el abatimiento del soporte adicional (5).

En un primer modo de funcionamiento, mostrado en la figura 5, el soporte adicional (5) está en la segunda posición y las cámaras (4) alojadas en las bahías (3) contrapuestas de la base (12) se disponen en posición nadiral. De esta forma se pueden captar simultáneamente panoramas completos (con una sombra cenital coincidente con la posición de la plataforma aérea) junto con la toma de imágenes estéreo verticales. En este primer modo de funcionamiento el tubo de conexión (8) está colocado en el segundo orificio (7.2).

Las cámaras (4) que se disponen en las bahías (3) de la base (12) se pueden colocar cada una de ellas con una determinada orientación o con una orientación girada 180º (en función de cómo se introduzca en la bahía).

En un segundo modo de funcionamiento el soporte adicional está también en la segunda posición y es similar al primer modo de funcionamiento pero rotado 180º de forma que las cámaras de la base (12) tengan posición cenital. Con este modo de funcionamiento se consiguen panoramas terrestres completos, salvo para la sombra proyectada en la posición de la cámara terrestre. La reconstrucción estereoscópica se puede hacer para las imágenes tomadas en posición cenital, como pueden ser por ejemplo techos y bóvedas de cualquier interior. En este segundo modo el tubo de conexión (8) está colocado en el segundo orificio (7.2).

Un tercer modo de funcionamiento, mostrado en las figuras 3 y 4, de la plataforma fotogramétrica es un modo mixto de pares estéreo múltiples y de estéreo-panoramas terrestres. En este modo de funcionamiento el soporte adicional (5) está en la primera posición y el tubo de conexión (8) está colocado en el primer orificio (7.1). De esta forma se consigue la reconstrucción estereoscópica múltiple de triple base de los elementos situados preferentemente en el plano acimutal. Mediante la rotación del tubo de conexión (6) se generan estéreopanoramas.

De nuevo, en función de cómo quede la orientación de las cámaras (4) contrapuestas que se alojan en las bahías (3) de la base (12) y la de la propia bahía (3) del soporte adicional (5) (0º o 180º), se podrán conformar pares estero con distintas baselíneas.

REIVINDICACIONES

1.- Plataforma fotogramétrica del tipo de las que permiten la instalación de una pluralidad de cámaras (4) en una estructura (1) de soporte con al menos unas paredes laterales (10) y una base (12) y que está caracterizada por que comprende:

-unos alojamientos en la estructura (1) de soporte en los que se encuentran unas bahías (3) destinadas a recibir las cámaras (4),

-al menos un soporte adicional (5) unido de forma abatible a la estructura (1) y que también comprende una bahía (3),

-un sistema de sujeción (6) destinado a fijar la inclinación del soporte adicional (5) respecto a la estructura (1),

-un elemento de conexión (7) dispuesto en la base (12) destinado a recibir un tubo de conexión (8) destinado a permitir su colocación en un dispositivo externo en el que se instala.

2.- Plataforma fotogramétrica según la reivindicación 1 caracterizada por que en la estructura (1) las caras laterales (10) y la base (12) están dispuestas ortogonalmente entre sí creando un espacio libre (11) entre ellas en el que está dispuesto el elemento de conexión (7).

3.- Plataforma fotogramétrica según la reivindicación 1 caracterizada por que en cada cara lateral (10) se encuentra una bahía (3) y en la base (12) se encuentran dos bahías (3) contrapuestas.

4.- Plataforma fotogramétrica según la reivindicación 1 caracterizada por que el soporte adicional (5) está unido a la base (12) dispuesto entre dos paredes laterales (10) y se desplaza entre una primera posición en la que está alineado con dicha base (12) y una segunda posición en la que queda ortogonal a la base (12) enfrentado a una tercera pared lateral (10).

5.- Plataforma fotogramétrica según la reivindicación 1 caracterizada por que el sistema de sujeción (6) comprende unos pasadores que fijan la posición del soporte adicional (5) respecto a la estructura (1).

6.- Plataforma fotogramétrica según la reivindicación 1 caracterizada por que comprende adicionalmente al menos una tapa (2) que se une a la estructura (1) de soporte en las bahías (3), destinada a asegurar la posición de las cámaras (4) en las bahías (3).

7.- Plataforma fotogramétrica según la reivindicación 1 caracterizada por que comprende adicionalmente una rosca de empalme (9) dispuesta en el extremo del tubo de conexión (8) destinada a su unión con el dispositivo externo en el que se instala.

8.- Plataforma fotogramétrica según la reivindicación 1 caracterizada por que el elemento de conexión (7) comprende un primer orificio (7.1) orientado hacia el soporte adicional (5) y un segundo orificio (7.2) orientado ortogonal al primer orificio (7.1) paralelo a la base (12).

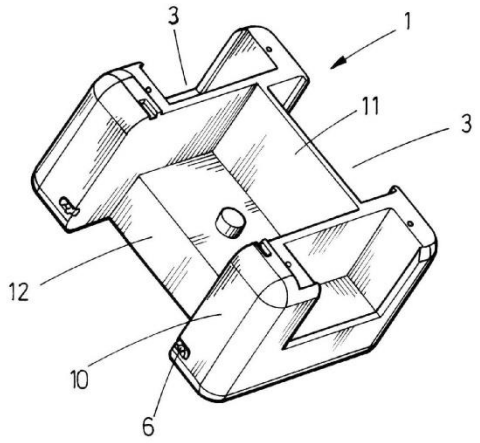


FIG.1

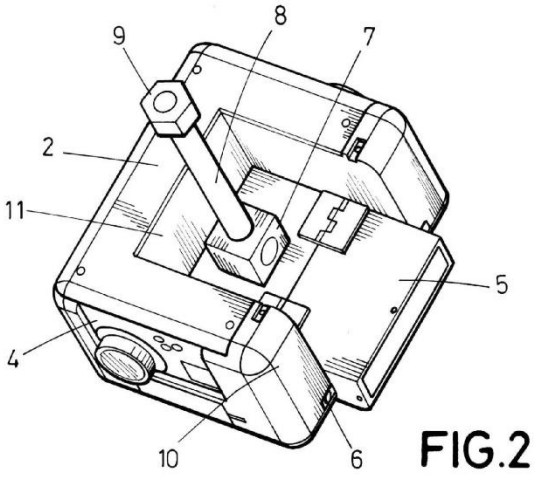


FIG.2

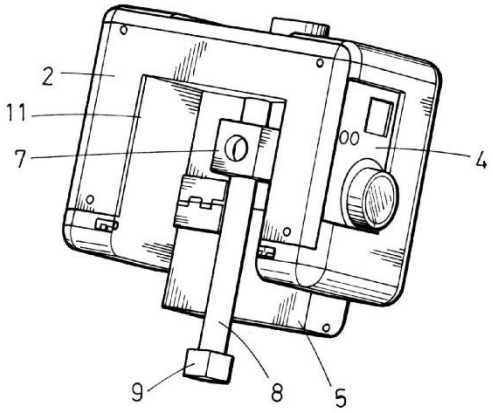


FIG.3

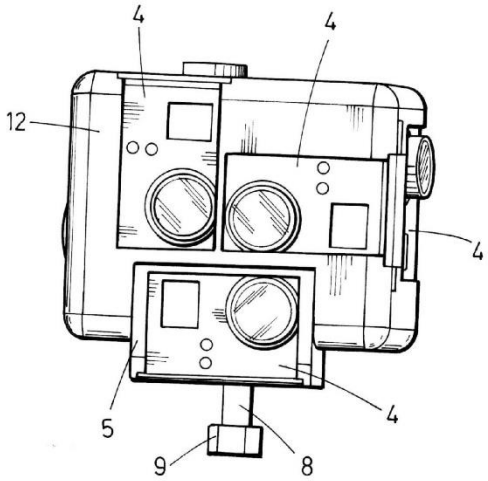


FIG.4

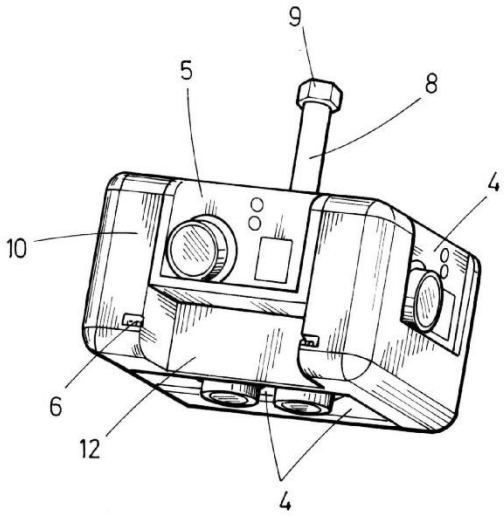


FIG.5

OFICINA ESPAÑOLA
DE PATENTES Y MARCAS

ESPAÑA

(21) N.º solicitud: 201331610

(22) Fecha de presentación de la solicitud: 04.11.2013

(32) Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TECNICA

(51) Int. Cl.: ***G03B17/56*** (2006.01)
G03B35/00 (2006.01)

DOCUMENTOS RELEVANTES

Categoría	(56) Documentos citados	Reivindicaciones afectadas
A	US 2004114038 A1 (LOUIS PIERRE) 17.06.2004, página 1, párrafo [1] – página 5, párrafo [74]; figuras 1-15.	1-8
A	US 3518929 A (GLENN WILLIAM E JR) 07.07.1970, columna 2, línea 50 – columna 5, línea 45; figuras 1-5.	1-8
A	WO 0206892 A2 (Z I IMAGING GMBH et al.) 24.01.2002, páginas 1-4; figuras 1-2.	1-8
A	ES 2357743 A1 (UNIV SALAMANCA) 29.04.2011, columna 2, línea 40 – columna 4, línea 31; figura 1.	1-8

Categoría de los documentos citados
X: de particular relevancia
Y: de particular relevancia combinado con otro/s de la misma categoría
A: refleja el estado de la técnica
O: referido a divulgación no escrita
P: publicado entre la fecha de prioridad y la de presentación de la solicitud
E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

☒ para todas las reivindicaciones ☐ para las reivindicaciones nº:

Fecha de realización del informe	Examinador	Página
17.12.2014	O. Fernández Iglesias	1/4

INFORME DEL ESTADO DE LA TÉCNICA

Nº de solicitud: 201331610

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G01C, F16M, G03B

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC

OPINIÓN ESCRITA

Nº de solicitud: 201331610

Fecha de Realización de la Opinión Escrita: 17.12.2014

Declaración

Novedad (Art. 6.1 LP 11/1986)	Reivindicaciones Reivindicaciones	1-8	**SI** **NO**
Actividad inventiva (Art. 8.1 LP11/1986)	Reivindicaciones Reivindicaciones	1-8	**SI** **NO**

Se considera que la solicitud cumple con el requisito de aplicación industrial. Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

Base de la Opinión.-

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como se publica.

OPINIÓN ESCRITA

Nº de solicitud: 201331610

1. Documentos considerados.-

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

Documento	Número Publicación o Identificación	Fecha Publicación
D01	US 2004114038 A1 (LOUIS PIERRE)	17.06.2004

2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de Patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración

El documento D01, al cual pertenecen las referencias que se indican a continuación, se considera el estado de la técnica más cercano a la invención tal y como se describe en la reivindicación 1. De la lectura del documento D01, y haciendo uso de la terminología de esta primera reivindicación de la solicitud, se puede apreciar que describe una plataforma fotogramétrica (3 -figura 3-) del tipo de las que permiten la instalación de una pluralidad de cámaras en una estructura de soporte (30 -figura 3-) con al menos unas paredes laterales (32 -figura 4-) y que comprende: unos alojamientos en la estructura de soporte en los que se encuentran unas bahías (80, 81 -figura 7-) destinadas a recibir las cámaras (4A, 4B, 4C, 4D -figura 5-); un elemento de conexión (82, 83 -figura 15-) dispuesto en la base destinado a recibir un tubo de conexión destinado a permitir su colocación en un dispositivo externo en el que se instala.

La invención reivindicada difiere del documento citado en que, en D01, no se describe un soporte adicional unido de forma abatible a la estructura y que comprende una bahía; además no existe un sistema de sujeción para fijar la inclinación del soporte adicional.

Por tanto, se deduce de los párrafos anteriores que ninguno de los documentos citados en este informe, ni ninguna combinación relevante de los mismos revela una plataforma fotogramétrica con las características descritas en la presente solicitud, y constituyen por tanto un reflejo del estado de la técnica. En consecuencia, la invención tal y como se recoge en las reivindicaciones 1 a 8 de la solicitud es nueva, se considera que implica actividad inventiva y que tiene aplicación industrial. (Artículos 6.1 y 8.1 de la Ley 11/86 de Patentes).

Patente 2: Plataforma fotogramétrica

MINISTERIO DE INDUSTRIA, ENERGIA Y TURISMO

Oficina Española de Patentes y Marcas

Nº SOLICITUD: **P201531242**
Nº PUBLICACIÓN: **ES2603657**
TITULAR/ES:
UNIVERSIDAD DE SALAMANCA
UNIVERSIDAD DE CASTILLA LA MANCHA

FECHA EXPEDICIÓN: 05/12/2017

TÍTULO
DE
PATENTE DE INVENCIÓN

Cumplidos los requisitos previstos en la vigente Ley 11/1986, de 20 de marzo, de Patentes, se expide el presente TÍTULO, acreditativo de la concesión de la Patente de Invención. La solicitud ha sido tramitada y concedida con realización del Informe sobre el Estado de la Técnica y **sin examen previo** de los requisitos sustantivos de patentabilidad.

Se otorga al titular un derecho de exclusiva en todo el territorio nacional, bajo las condiciones y con las limitaciones previstas en la Ley de Patentes. La duración de la patente será de **veinte años** contados a partir de la fecha de presentación de la solicitud (31/08/2015).

La patente se concede sin perjuicio de tercero y sin garantía del Estado en cuanto a la validez y a la utilidad del objeto sobre el que recae.

Para mantener en vigor la patente concedida, deberán abonarse las tasas anuales establecidas, que se pagarán por años adelantados. Asimismo, deberá explotarse el objeto de la invención, bien por su titular o por medio de persona autorizada de acuerdo con el sistema de licencias previsto legalmente, dentro del plazo de cuatro años a partir de la fecha presentación de la solicitud de patente, o de tres años desde la publicación de la concesión en el Boletín Oficial de la Propiedad Industrial, aplicándose el plazo que expire más tarde.

OFICINA ESPAÑOLA DE PATENTES Y MARCAS

Fdo.: Ana María Redondo Mínguez
Jefe/a de Servicio de Actuaciones Administrativas
(P.D. del Director del Departamento de Patentes e I.T., resolución 05/09/2007)

(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 603 657**

(21) Número de solicitud: 201531242

(51) Int. CI.:

G03B 35/08 (2006.01)
G12B 9/08 (2006.01)

(12) PATENTE DE INVENCIÓN B1

(22) Fecha de presentación:
31.08.2015

(43) Fecha de publicación de la solicitud:
28.02.2017

Fecha de concesión:
05.12.2017

(45) Fecha de publicación de la concesión:
14.12.2017

(73) Titular/es:
UNIVERSIDAD DE SALAMANCA (75.0%)
PATIO DE ESCUELAS, 1
37008 SALAMANCA (Salamanca) ES y
UNIVERSIDAD DE CASTILLA LA MANCHA (25.0%)

(72) Inventor/es:
GONZALEZ AGUILERA, Diego;
RODRIGUEZ GONZALVEZ, Pablo;
GESTO DÍAZ , Manuel y
HERNÁNDEZ LÓPEZ, David

(74) Agente/Representante:
PONS ARIÑO, Ángel

(54) Título: **DISPOSITIVO AUTÓNOMO DE GENERACIÓN DE MODELOS FACIALES EN TRES DIMENSIONES**

(57) Resumen:
Dispositivo autónomo de generación de modelos faciales en tres dimensiones.
La presente invención es un dispositivo autónomo de generación de modelos faciales que permite generar un modelo facial en tres dimensiones, para almacenado en una base de datos, ser analizado y comparado con otros modelos faciales en tres dimensiones con fines preferiblemente policiales. Para ello, este dispositivo comprende un cuerpo portátil que a su vez comprende al menos dos captadores de imagen y rango, un mango, una unidad de procesado de datos una unidad de representación gráfica y de control y una batería.

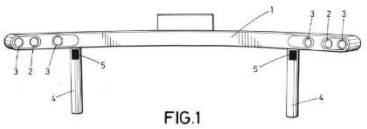


FIG.1

ES 2 603 657 B1

DISPOSITIVO AUTÓNOMO DE GENERACIÓN DE MODELOS FACIALES EN TRES DIMENSIONES

DESCRIPCIÓN

OBJETO DE LA INVENCIÓN

El objeto de la presente invención es un dispositivo autónomo de generación de modelos faciales en tres dimensiones.

Más concretamente, la presente invención permite generar un modelo facial en tres dimensiones, almacenarlo en una base de datos para analizarlo y compararlo con otros modelos faciales en tres dimensiones con fines preferiblemente policiales.

ANTECEDENTES DE LA INVENCIÓN

Actualmente, son conocidos dispositivos para la reconstrucción y el reconocimiento facial con fines policiales, tal como cámaras fotográficas convencionales, para tomar las imágenes que se requieren para realizar la ficha de un individuo.

Más concretamente, dichas fotografías son tomadas en una serie de posiciones predefinidas (frontal, perfil derecho y semi-perfil izquierdo) con lo que en ocasiones resulta difícil identificar a individuos que son fotografiados o grabados desde perspectivas diferentes, siendo realmente complicada su identificación no sólo por sistemas automáticos sino por el propio personal experto de la policía.

Asimismo, en otras ocasiones el individuo modifica su aspecto facial, incluso con cirugía estética, siendo imposible la identificación del mismo mediante las técnicas fotográficas actualmente utilizadas por la policía. Esto es debido a que las imágenes en dos dimensiones tomadas actualmente no son métricas, y por tanto no se obtienen los patrones faciales biométricos clave para la identificación de la persona en esta situación.

Actualmente, para solventar este problema existen soluciones de software comercial para el reconocimiento facial en dos o tres dimensiones. La problemática de estos sistemas

radica en que los modelos faciales en tres dimensiones son generados a partir de las tres imágenes en dos dimensiones citadas anteriormente, con lo que únicamente se consiguen distintas perspectivas (picado, contrapicado, etc.) de las tres imágenes en dos dimensiones. De este modo, no es posible obtener un modelo facial en tres dimensiones suficientemente acorde con la realidad para mostrar los patrones faciales biométricos del individuo.

Por otro lado, en otros sectores ajenos a la investigación policial, se conocen dispositivos para la generación de modelos faciales en tres dimensiones.

Por ejemplo en el sector de la cirugía estética, se conocen dispositivos que realizan modelos en tres dimensiones para mostrar al paciente el posible resultado después de la cirugía, ayudando en muchos casos a tomar la decisión estética deseada. Estos dispositivos son capaces de realizar modelos fáciles en tres dimensiones utilizando costosos y pesados escáneres que requieren de un software instalado en un ordenador externo al dispositivo, y debido a su amplio consumo energético tiene que estar conectados a la red eléctrica para poderlos utilizar.

Otro sector en que se realizan modelos faciales en tres dimensiones es en los videojuegos. Actualmente, es conocido un sistema de reconstrucción facial basado en cámaras que captan el espectro visible. Más concretamente, este sistema está compuesto por un elevado número de cámaras fotográficas réflex y múltiples fuentes de luz dispuestas en un armazón metálico fijo. La captura no se realiza en tiempo real, sino que son necesarias varias tomas con diferentes condiciones de iluminación que deben ser postprocesadas mediante un ordenador externo. Mediante este sistema se obtiene un modelo en tres dimensiones de gran calidad pero su configuración imposibilita el desplazamiento del equipo, requiere de fuentes de luz externa y depende de un sistema de procesamiento externo para su funcionamiento.

DESCRIPCIÓN DE LA INVENCIÓN

La presente invención describe un dispositivo autónomo de generación de modelos faciales en tres dimensiones, que comprende:

- un cuerpo portátil horizontalmente alargado,

- al menos dos captadores de imagen y rango alineados entre sí, estando cada captador emplazado en los extremos opuestos de la misma superficie posterior del cuerpo para generar al menos una imagen en tres dimensiones de la cara de un individuo,
- al menos un mango ergonómico con dos extremos, siendo uno de los extremos insertable en el cuerpo portátil, y el otro extremo susceptible de ser sujetado,
- un botón de disparo de los sensores de imagen situado en el exterior del mango destinado a ser pulsado por el usuario, preferentemente personal policial, para accionar el captador de imagen y rango,
- una unidad de procesado de datos emplazada en el interior del cuerpo portátil, y vinculada con los sensores de imagen y rango, y con el botón de disparo, para generar el modelo facial en tres dimensiones,
- una unidad de representación gráfica y de control emplazada en la superficie anterior del cuerpo portátil, que está vinculada con la unidad de procesado de datos para representar visualmente la captura de los sensores de imágenes, de los modelos faciales en tres dimensiones, y para controlar la generación del modelo facial en tiempo real, y
- una batería emplazada en el interior del cuerpo portátil para alimentar eléctricamente los componentes electrónicos del dispositivo.

Preferentemente, el cuerpo portátil es simétrico y horizontalmente curvado para obtener una mayor superficie lateral de la cara del individuo.

Más concretamente, cada captador de imagen y rango comprende una cámara RGB y un sensor de profundidad para la obtención, en una única toma, de la imagen facial completa del individuo. Este sensor de profundidad comprende preferentemente un emisor y un receptor de infrarrojos, y presenta una velocidad de captura elevada de modo que los posibles movimientos del usuario no afectan a la toma. Adicionalmente, al utilizar un sensor de profundidad con una señal cuya longitud de onda es no visible, la toma no requiere de unas condiciones de iluminación especiales, y por tanto puede ser utilizada en cualquier situación lumínica o ambiental.

Concretamente, la unidad de procesado de datos comprende una estructura tipo raspberry pi con software que permite realizar el modelo facial en tres dimensiones

integrado.

Preferentemente, la unidad de procesado de datos comprende un módulo de comunicación inalámbrica para recibir y trasmitir datos con un dispositivo exterior, tal como una base de datos que preferentemente se encuentra en la nube. Adicionalmente, esta base de datos de modelos faciales en tres dimensiones permite comparar los parámetros faciales biométricos en tiempo real de individuos previamente identificados por la policía o autoridad similares.

Adicionalmente, también el cuerpo portátil comprende un módulo de transferencia de datos, preferiblemente un puerto USB, vinculado a la unidad de procesado de datos y a la batería. Este puerto USB permite la carga de la batería, así como recibir y trasmitir datos con una memoria USB o similar.

La unidad de representación gráfica y de control es una pantalla táctil que permite ver y controlar la generación del modelo facial.

Otra de las ventajas de este dispositivo es que el mango es desmontable y por tanto se facilita su transporte.

De este modo la presente invención es de especial interés con fines policiales debido a que es un dispositivo autónomo, portátil y ligero que puede realizar un modelo facial en tres dimensiones con malas condiciones de iluminación y enviarlo para ser comparado en una base de datos de forma inalámbrica, permitiendo una identificación y reconocimiento facial inmediata, fiable y eficiente.

DESCRIPCIÓN DE LOS DIBUJOS

Para complementar la descripción que se está realizando y con objeto de ayudar a una mejor comprensión de las características de la invención, de acuerdo con un ejemplo preferente de realización práctica de la misma, se acompaña como parte integrante de dicha descripción, un juego de dibujos en donde con carácter ilustrativo y no limitativo, se ha representado lo siguiente:

Figura 1.- Muestra una vista esquemática de la superficie posterior del dispositivo autónomo de generación de modelos faciales en tres dimensiones.

Figura 2.- Muestra una vista esquemática de la superficie anterior del dispositivo autónomo de generación de modelos faciales en tres dimensiones.

REALIZACIÓN PREFERENTE DE LA INVENCIÓN

En una realización preferente de esta invención, tal y como se muestra en la figura 1 y 2, el dispositivo autónomo de generación de modelos faciales en tres dimensiones con fines policiales comprende un cuerpo portátil (1), que tiene una configuración simétrica horizontalmente alargada y curvada que resulta optimizada para el proceso del modelo facial en tres dimensiones del individuo.

Más concretamente, en cada extremo de la superficie posterior, es decir la superficie que apunta a la cara del individuo cuando se realiza el modelo, del cuerpo portátil (1) comprende un captador de imagen y rango compuestos por una cámara RGB (2) y un sensor de profundidad (3) que comprende un emisor de infrarrojos y un receptor de infrarrojos. Dichos captadores de imagen están sincronizados entre sí y emplean longitudes de onda no visibles.

Preferentemente, el cuerpo portátil (1) comprende en su cara inferior dos mangos (4) ergonómicos desmontables para facilitar el manejo del cuerpo portátil (1) por el personal policial. Cada mango (4) comprende un botón de disparo (5) en su extremo superior interior, es decir en su extremo más próximo al cuerpo portátil (1) y en su superficie posterior. De este modo se facilita la captura de la imagen para realizar el modelo en tres dimensiones en el momento deseado por el personal policial.

Para que el personal policial pueda supervisar en tiempo real el modelo el cuerpo portátil (1) dispone de una unidad de representación gráfica y de control, tal como una pantalla táctil (7) en su superficie anterior, es decir la superficie que apunta al personal de policía cuando se realiza el modelo. Concretamente en su parte central. De este modo, únicamente el personal policial puede visualizar la toma de la imagen y la generación del modelo en tres dimensiones.

Concretamente, en el interior del cuerpo portátil (1) comprende una unidad de procesado de datos, preferente una estructura rasberry pi (6), que mediante un software integrado transforma la información capturada por la cámara RGB (2) y el sensor de profundidad (3) para generar el modelo facial en tiempo real.

Conjuntamente, con dicha estructura rasberry pi (6), está incluido un módulo de comunicación inalámbrica (8) que permite la transferencia de los modelos finales a un sistema de almacenamiento externo preferiblemente a una base de datos que se encuentra en la nube. De este, modo si el individuo ha sido previamente identificado por la policía, o una autoridad similar, el personal policial puede conocerlo en tiempo real.

Preferentemente, en el interior del cuerpo portátil (1) se encuentra una batería (10) para permitir el funcionamiento autónomo del dispositivo.

Adicionalmente, el cuerpo portátil (1) también comprende un módulo de transferencia de datos, tal como un puerto USB (9), para descargar datos y recarga la batería (10).

En esta realización preferente, el peso completo del dispositivo es aproximadamente 1.3 kilógramos y presenta una autonomía mínima de 2 horas.

REIVINDICACIONES

1.- Dispositivo autónomo de generación de modelos faciales en tres dimensiones, caracterizado porque comprende:

- un cuerpo portátil (1) horizontalmente alargado,
- al menos dos captadores de imagen y rango alineados entre sí, estando cada captador emplazado en los extremos opuestos de la misma superficie posterior del cuerpo portátil (1) para generar al menos una imagen en tres dimensiones de la cara de un individuo,
- al menos un mango (4) ergonómico con dos extremos, siendo uno de los extremos insertable en el cuerpo portátil (1), y el otro extremo susceptible de ser sujetado,
- un botón de disparo (5) de los sensores de imagen situado en el exterior del mango (4) destinado a ser pulsado por el usuario para accionar el captadores de imagen y rango,
- una unidad de procesado de datos emplazada en el interior del cuerpo portátil (1), y vinculada con el sensor de imagen y rango con el botón de disparo (5), para generar el modelo facial en tres dimensiones,
- una unidad de representación gráfica y de control emplazada en la superficie anterior del cuerpo portátil (1), que está vinculada con la unidad de procesado de datos para representar visualmente la captura de los sensores de imágenes, de los modelos faciales en tres dimensiones, y para controlar la generación del modelo facial en tiempo real, y
- una batería (10) emplazada en el interior del cuerpo portátil (1) para alimentar eléctricamente los componentes electrónicos del dispositivo.

2.- Dispositivo según la reivindicación 1, caracterizado porque el cuerpo portátil (1) es simétrico y horizontalmente curvado.

3.- Dispositivo según la reivindicación 1, caracterizado porque cada captador de imagen y rango comprenden una cámara RGB (2) y un sensor de profundidad (3) para la obtención en una única toma de la imagen facial completa.

4.- Dispositivo según la reivindicación 1, caracterizado porque la unidad de procesado de

datos comprende una estructura tipo rasberry pi (6) con software que permite el modelo facial en tres dimensiones integrado.

5.- Dispositivo según la reivindicación 1, caracterizado porque la unidad de procesado de datos comprende un módulo de comunicación inalámbrica (8) para recibir y trasmitir datos con un dispositivo exterior, tal como una base de datos.

6.- Dispositivo según la reivindicación 1, caracterizado porque el cuerpo portátil (1) comprende un módulo de transferencia de datos vinculado a la unidad de procesado de datos y a la batería (10).

7.- Dispositivo según la reivindicación 6, caracterizado porque el módulo de transferencia de datos es un puerto USB (9).

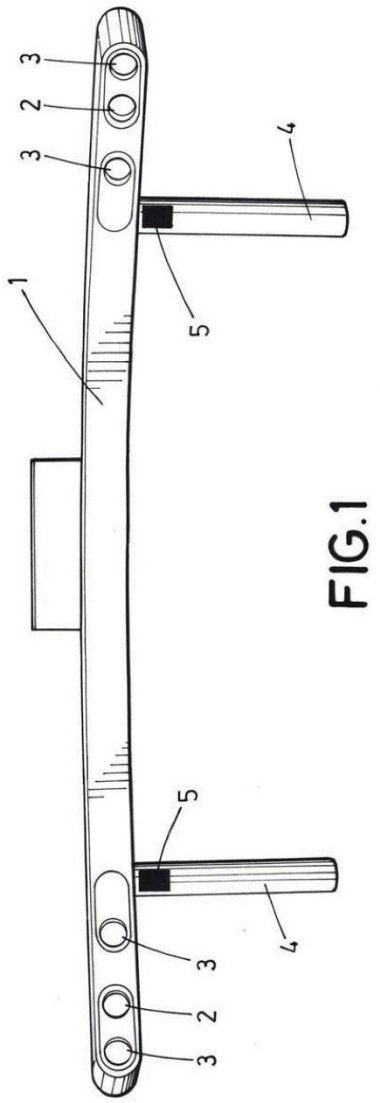


FIG.1

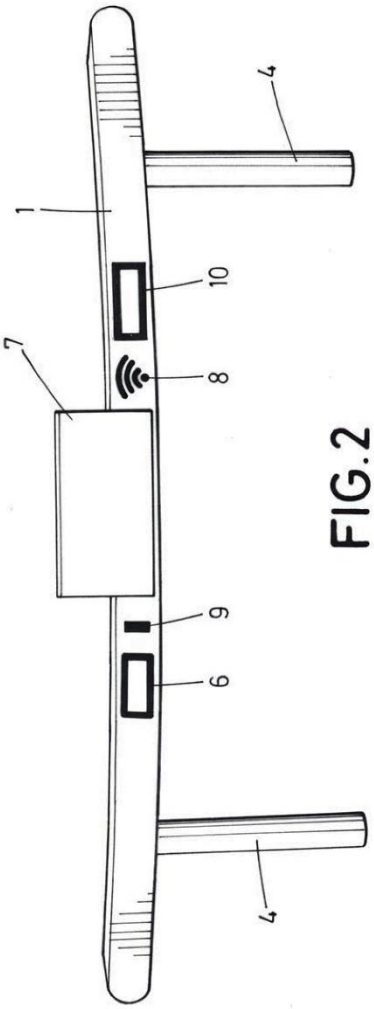


FIG.2

OFICINA ESPAÑOLA
DE PATENTES Y MARCAS

ESPAÑA

(21) N.º solicitud: 201531242

(22) Fecha de presentación de la solicitud: 31.08.2015

(32) Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TECNICA

(51) **Int. Cl.:** ***G03B35/08*** (2006.01)
G12B9/08 (2006.01)

DOCUMENTOS RELEVANTES

Categoría	(56) Documentos citados	Reivindicaciones afectadas
Y	US 2011242286 A1 (PACE VINCENT et al.) 06/10/2011, párrafos [24 - 30]; párrafos [34 - 36]; párrafos [45 - 69]; figuras 1 - 7.	1-7
Y	US 2006204239 A1 (INABA MINORU) 14/09/2006, Columna 2, línea 56 - columna 3, línea 50; columna 4, línea 27 - columna 5, línea 5; columna 7, línea 57 - columna 9, línea 19; figuras 1 - 2.	1-7
A	US 2010239240 A1 (CAMERON JAMES et al.) 23/09/2010, Párrafos [21 - 29]; párrafos [36 - 58]; figuras 1 - 7.	1-7
A	US 2015144759 A1 (CHANG YU-CHENG) 28/05/2015, Figuras 1 - 2. párrafos [5 - 6];	2
A	Stereo depth perception with Raspberry Pi. [en línea] Noviembre 2014 [Recuperado el 31-10-16] Recuperado de Internet: https://web.archive.org/web/20141109193541/http://makezine.com/2014/11/03/stereo-depth-perception-with-raspberry-pi/	4
A	US 4431290 A (KENNEDY JOHN H) 14/02/1984, Columna 1, líneas 8 - 16; líneas 57 - 64; columna 2, líneas 15 - 22; columna 3, líneas 23 - 35; columna 3, Línea 55 - columna 4, línea 10; figuras 1 - 5.	1-7

Categoría de los documentos citados
X: de particular relevancia
Y: de particular relevancia combinado con otro/s de la misma categoría
A: refleja el estado de la técnica
O: referido a divulgación no escrita
P: publicado entre la fecha de prioridad y la de presentación de la solicitud
E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado
☒ para todas las reivindicaciones ☐ para las reivindicaciones nº:

Fecha de realización del informe	Examinador	Página
02.11.2016	J. M. Vazquez Burgos	1/5

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G03B, G12B

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI, INTERNET

OPINIÓN ESCRITA

Nº de solicitud: 201531242

Fecha de Realización de la Opinión Escrita: 02.11.2016

Declaración

Novedad (Art. 6.1 LP 11/1986)	Reivindicaciones	1-7	**SI**
	Reivindicaciones		**NO**
Actividad inventiva (Art. 8.1 LP11/1986)	Reivindicaciones		**SI**
	Reivindicaciones	1-7	**NO**

Se considera que la solicitud cumple con el requisito de aplicación industrial. Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

Base de la Opinión.-

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como se publica.

1. Documentos considerados.-

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

Documento	Número Publicación o Identificación	Fecha Publicación
D01	US 2011242286 A1 (PACE VINCENT et al.)	06.10.2011
D02	US 2006204239 A1 (INABA MINORU)	14.09.2006
D03	US 2010239240 A1 (CAMERON JAMES et al.)	23.09.2010
D04	US 2015144759 A1 (CHANG YU-CHENG)	28.05.2015
D05	Stereo depth perception with Raspberry Pi. [en línea] Noviembre 2014 [Recuperado el 31-10-16]Recuperado de Internet: https://web.archive.org/web/20141109193541/http://makezine.com/2014/11/03/stereo-depth-perception-with-raspberry-pi/	09.11.2014

2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de Patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración

El documento del estado de la técnica más próximo a la invención es D01 y divulga un método y un sistema para la toma de imágenes estereoscópicas (3D), ya sea de vídeo o estáticas, mediante sendas cámaras 2D y posterior procesamiento de sus imágenes. El documento D01 incorpora por referencia (párrafos 34, 53) al documento D03, cuyo contenido se considera parte de D01, en el cual se precisa que las cámaras pueden situarse en una plataforma ajustable.

Reivindicación 1

Para mayor claridad en la diferencias entre la invención reivindicada en 1 y el documento D01 del estado de la técnica más próximo, se reproduce seguidamente el texto de dicha reivindicación, despojado de sus referencias originales e introduciendo donde proceda las del documento D01. Asimismo, aquellas partes del texto que pudieran no estar incluidas en D01 se señalarían entre corchetes y en negrita.

Dispositivo [**autónomo**] de generación de modelos faciales en tres dimensiones (párrafo 24), caracterizado porque comprende:

- un cuerpo portátil (550 en D03; párrafos 37-39 en D03) horizontalmente alargado,
- al menos dos captadores de imagen y rango alineados entre sí (612L, 612R, 636, 638, 680; 510L, 510R en D03; párrafos 28, 45-46, 51-52, 63), estando cada captador emplazado en [**los extremos opuestos de la misma superficie posterior d**]el cuerpo portátil (550 en D03) para generar al menos una imagen en tres dimensiones de la cara de un individuo (122),
- [**al menos un mango ergonómico con dos extremos, siendo uno de los extremos insertable en el cuerpo portátil, y el otro extremo susceptible de ser sujetado,**]
- un botón de disparo de los sensores de imagen [**situado en el exterior del mango**] destinado a ser pulsado por el usuario para accionar el captadores de imagen y rango (768; párrafo 60),
- una unidad de procesado de datos (752; 664 en D03) emplazada en el interior del cuerpo portátil (párrafo 45 en D03), y vinculada con el sensor de imagen y rango con el botón de disparo (párrafos 54, 60; figura 7), para generar el modelo facial en tres dimensiones,
- una unidad de representación gráfica y de control (772) emplazada en [**la superficie anterior d**]el cuerpo portátil (párrafo 45 en D03), que está vinculada con la unidad de procesado de datos para representar visualmente la captura de los sensores de imágenes, de los modelos faciales en tres dimensiones, y para controlar la generación del modelo facial en tiempo real (párrafos 57, 69), y
- [**una batería emplazada en el interior del cuerpo portátil para alimentar eléctricamente los componentes electrónicos del dispositivo.**]

Se considera que, al estar las cámaras separadas una distancia determinada, el soporte (plataforma) donde se ubican forzosamente habrá de tener una cierta dimensión en el sentido horizontal, lo que implícitamente supone que el mismo ha de ser alargado en sentido horizontal.

Aunque el documento D01 no especifica explícitamente la existencia de un botón de disparo, sí que menciona en su párrafo 60 una interfaz de operación de la cámara por un operador, en la que un experto en la materia no necesitaría de actividad inventiva para incluir el mencionado mando de disparo. Por otro lado, D01 tampoco menciona explícitamente una unidad de visualización, sino tan solo una de procesado gráfico 772. No obstante, la conexión de una unidad visualización una vez disponibles las señales gráficas de entrada (775) resulta también un problema de obvia resolución para un experto en la materia, tanto más cuanto que D01 contempla la posibilidad de conectar visualizadores (762) a la unidad de control.

OPINIÓN ESCRITA

Nº de solicitud: 201531242

Tenidas en cuenta las precisiones anteriores, las diferencias entre D01 y el invención reivindicada en 1 serían:

a) a D01 (a través de D03) contempla la ubicación de las cámaras en una plataforma que permita situarlas en la orientación deseada, pero no concreta la forma de dicha plataforma, ni que esta disponga de un mango. En cualquier caso, la orientación de dichas cámaras puede ser de manera que sus ejes sean paralelos o converjan en un punto (párrafo 50; figura 1).
b) D01 no concreta tampoco en qué parte de la plataforma se ubicarían las cámaras, el botón de disparo o la unidad de representación gráfica.
c) D01 no detalla la existencia de una batería que alimente el dispositivo.

Estas diferencias devienen en los siguientes efectos técnicos:

a) Se evita apoyar el soporte en las manos del operario que lo maneja, reduciendo la sensibilidad del dispositivo a movimientos o vibraciones de aquellas.
b) Las dimensiones del soporte se reducen al mínimo, al no sobrar material más allá de la ubicación de las cámaras (que se entiende es fija).
c) Es posible acceder a la interfaz gráfica sin necesidad de interrumpir el funcionamiento del dispositivo (por no disponerse el operario en la línea de visión de las cámaras).
d) El pulsado de los botones de disparo no transmite vibraciones directas al soporte.
e) El dispositivo dispone de alimentación autónoma.

Por tanto, de todo ello se deduce que los problemas técnicos objetivos a resolver son:

a) Diseñar un soporte horizontal de las cámaras que evite el contacto directo con las manos del operario que lo maneja.
b) Diseñar dicho soporte con la longitud horizontal mínima (asumiendo siempre una misma separación entre cámaras).
c) Disponer los elementos de disparo y de interfaz gráfica de manera que los primeros puedan pulsarse mientras el operador sostiene el dispositivo, y la segunda pueda operarse sin necesidad de interrumpir el funcionamiento del dispositivo.
d) Dotar al dispositivo de alimentación independiente de la de red.

De estos problemas, el b) es de resolución evidente para un experto en la materia, mientras que, con respecto al resto, el documento D02 muestra un dispositivo para la fotografía estereoscópica, donde las cámaras se montan en una estructura dotada de al menos un mango (14, 15) para portarla por un operario sin necesidad de que este entre en contacto con la base sobre la que descansan las cámaras. Además dispone de visualizadores y batería (columna 8, líneas 21-38), así como de disparadores en uno de los mangos (18; columna 8, líneas 45-53). A partir de las enseñanzas de este documento en cuanto a la construcción de un soporte para cámaras de fotografía en 3D y ubicación en él de los elementos de alimentación, visualizadores y de disparo, un experto en la materia, en combinación con D01 resolvería los problemas técnicos objetivos sin necesidad del recurso a la actividad inventiva.

Por lo tanto a la luz de la combinación de D01 con D02, la invención reivindicada en 1 carece de actividad inventiva tal como se establece dicho requisito en el artículo 8 de la Ley de Patentes de 1986.

Reivindicaciones 2 a 7

La forma reivindicada para el soporte en 2 se derivaría de forma evidente para un experto en la materia a partir de la posibilidad comprendida en D01, de que los ejes de las cámaras puedan converger en un punto (figura 1; párrafo 25). En este sentido, la solución de la orientación de las cámaras se ajuste a esta condición gracias a un montaje solidario con el soporte constituye una realización particular del problema técnico de cómo montar las cámaras para que sus ejes converjan en un punto, obvia para dicho experto. El documento D04 ilustra como un montaje de este tipo en una cámara forma parte del estado de la técnica (figura 1; párrafos 5-6).

Las cámaras y el sensor de profundidad reivindicado en 3 estarían comprendidos en D01 (párrafos 20, 28).

Los medios específicos de procesamiento reivindicados en 4 constituyen una realización particular del dispositivo de procesamiento mencionado en D01 (figura 7), obvio para un experto en la materia. El documento D05 constituye un ejemplo de aplicación de los mismos a este sector de la técnica.

La comunicación inalámbrica reivindicada en 5 está incluida en D01 (párrafos 61-62), lo mismo que el módulo de transferencia de datos (párrafo 61) reivindicado en 6, y la conexión USB objeto de 7 (párrafo 62).

En consecuencia, a la luz de la combinación de D01 con D02, y habiendo tenido en cuenta las correspondientes relaciones de dependencia, se concluye que las invenciones reivindicadas en 2 a 7 no poseen actividad inventiva, conforme dicho requisito se define en el artículo 8 de la Ley de Patentes de 1986.

Patente 3: BRazo Automatizado MUltisensor para la Reconstrucción 3D (BRAMUR3D)

MINISTERIO DE INDUSTRIA, ENERGIA Y TURISMO

Oficina Española de Patentes y Marcas

Nº SOLICITUD: **P201530194**
Nº PUBLICACIÓN: **ES2580128**
TITULAR/ES:
UNIVERSIDAD DE SALAMANCA
UNIVERSIDAD DE CASTILLA LA MANCHA

FECHA EXPEDICIÓN: 25/05/2017

TÍTULO
DE
PATENTE DE INVENCIÓN

Cumplidos los requisitos previstos en la vigente Ley 11/1986, de 20 de marzo, de Patentes, se expide el presente TÍTULO, acreditativo de la concesión de la Patente de Invención. La solicitud ha sido tramitada y concedida con realización del Informe sobre el Estado de la Técnica y **sin examen previo** de los requisitos sustantivos de patentabilidad.

Se otorga al titular un derecho de exclusiva en todo el territorio nacional, bajo las condiciones y con las limitaciones previstas en la Ley de Patentes. La duración de la patente será de **veinte años** contados a partir de la fecha de presentación de la solicitud (18/02/2015).

La patente se concede sin perjuicio de tercero y sin garantía del Estado en cuanto a la validez y a la utilidad del objeto sobre el que recae.

Para mantener en vigor la patente concedida, deberán abonarse las tasas anuales establecidas, que se pagarán por años adelantados. Asimismo, deberá explotarse el objeto de la invención, bien por su titular o por medio de persona autorizada de acuerdo con el sistema de licencias previsto legalmente, dentro del plazo de cuatro años a partir de la fecha presentación de la solicitud de patente, o de tres años desde la publicación de la concesión en el Boletín Oficial de la Propiedad Industrial, aplicándose el plazo que expire más tarde.

OFICINA ESPAÑOLA DE PATENTES Y MARCAS

Fdo.: Ana María Redondo Mínguez
Jefe/a de Servicio de Actuaciones Administrativas
(P.D. del Director del Departamento de Patentes e I.T., resolución 05/09/2007)

VNiVERSiDAD D SALAMANCA

OFICINA DE TRANSFERENCIA DE RESULTADOS DE INVESTIGACIÓN (OTRI)

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

IMPRESO DE SOLICITUD DE PATENTE EN LA USAL

Persona de contacto durante la tramitación de la solicitud:

Nombre y apellidos: Diego González Aguilera

Departamento y Centro: Departamento de Ingeniería Cartográfica y del Terreno

Tfno: 625399698 e.mail: daguilera@usal.es

1.- TÍTULO DE LA INVENCIÓN:

BRazo Automatizado MUltisensor para la Reconstrucción 3D (BRAMUR3D)

2.- DATOS DE LOS SOLICITANTES:

Inventores de la Universidad de Salamanca:

Apellidos y nombre	Categoría académica	Dpto/Centro	% participación
1. Manuel Gesto Díaz	PAS	Ingeniería Cartográfica y del Terreno	30%
2. Pablo Rodríguez Gonzálvez	PI	Ingeniería Cartográfica y del Terreno	20%
3. Diego González Aguilera	P.TU	Ingeniería Cartográfica y del Terreno	20%
4. Jesús Fernández Hernández	P. Asociado	Ingeniería Cartográfica y del Terreno	20%
	Total % de participación de la USAL		90%

Colaboradores pertenecientes a otras entidades:

Apellidos y nombre	DNI	Entidad	Categoría profesional
5. Hernández López, David	07555547R	IDR/UCLM	P.TU

Existe convenio de copropiedad entre las partes: ☐**Sí** ☒**No**

Firma de todos los inventores **Fecha: 21/10/2013**

1.	**2.**	**3.**
4.	**5.**	**6.**

EVALUACION DE LA PATENTABILIDAD

VNiVERSiDAD D SALAMANCA

OFICINA DE TRANSFERENCIA DE RESULTADOS DE INVESTIGACIÓN (OTRI)

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

Los datos que se solicitan a continuación son meramente orientativos para la OTRI sobre la patentabilidad de la invención y posibilitan la agilización de los trámites. Trate de responder a los apartados que pueda y deje en blanco los que no sepa responder para rellenarlos junto con el personal técnico de la OTRI.

3.- OBJETO DE LA INVENCIÓN

Defina (por favor), su resultado (elija una o varias opciones):

- ☒ un nuevo producto (considerar 'producto' en sentido general)
- ☐ un nuevo procedimiento de invención
- ☐ mejora de un producto existente
- ☐ mejora de un proceso existente
- ☐ una idea
- ☐ un servicio nuevo o mejorado

4.- ORIGEN DE LA INVENCIÓN

La invención es resultado de :

- ☒ un proyecto de investigación
- ☐ un contrato de investigación (*citar nombre de la empresa*)
- ☐ un proyecto de Investigación con participación empresarial (*citar código de referencia, entidad financiadora y empresa*)
- ☐ otros

5.- DESCRIPCIÓN GENERAL DE LA INVENCIÓN

Describir brevemente (máximo 300 palabras) el objeto de la invención, en qué consiste, qué problema técnico resuelve y qué ventajas aporta respecto al estado de la técnica actual.

RESUMEN

El Brazo Automatizado Multisensor para la Reconstrucción 3D (BRAMUR3D) es un soporte en forma de tijera (Figura1) diseñado para permitir la reconstrucción tridimensional empleando diferentes tipos de sensores, tanto activos (láser) como pasivos (cámaras).

VNiVERSiDAD D SALAMANCA

OFICINA DE TRANSFERENCIA DE RESULTADOS DE INVESTIGACIÓN (OTRI)

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

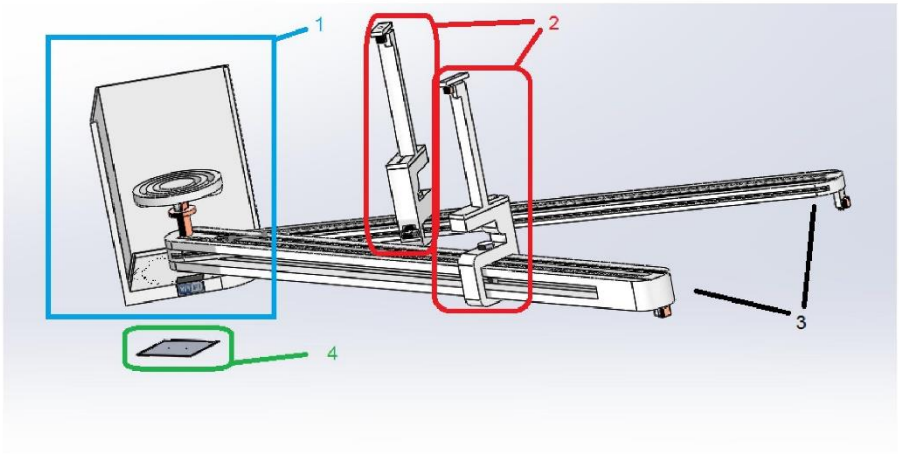


Figura 1. Diseño CAD renderizado de la invención BRAMUR3D.

La reconstrucción e ingeniería inversa de piezas industriales requiere gran detalle y precisión. Para conseguir esta reconstrucción existen diferentes tipos de sensores: activos, pasivos y combinación de los anteriores. En cualquier caso, para capturar el objeto de forma completa es necesario que los sensores accedan a la totalidad de la superficie del objeto, lo que se puede conseguir bien manteniendo estático el objeto y desplazando los sensores alrededor del objeto, o bien manteniendo estáticos los sensores y moviendo el propio objeto con respecto a los mismos, siendo ideal un sistema, como el aquí propuesto, que incorpora simultáneamente las dos alternativas. En definitiva, se plantea la necesidad de disponer de información precisa sobre el posicionamiento relativo del sensor o sensores con respecto al objeto en todo el proceso, a fin de que las diferentes capturas puedan ser fusionadas para derivar el modelo tridimensional completo final del objeto.

Se ha diseñado un soporte en forma de tijera que incorpora un sistema de amarre y posicionamiento de sensores para poder embarcar cualquier tipo de sensor y lograr una posición espacial del mismo con ciertos grados de libertad. Más concretamente, se incorporan 2 brazos (3), uno para cada sensor, los cuales a través de un sistema de movimiento de precisión en su extremo son capaces de conseguir el ángulo de toma deseado. Estos brazos (3) también incorporan un sistema de comunicación y railes para cada uno de los soportes de los sensores (2). Estos soportes (2), además de tener un sistema de movimiento propio que permite fijar la distancia o alejamiento al objeto a reconstruir, incorporan unos carriles con los que se puede transmitir la información desde la unidad de control a cada uno de los soportes sobre los brazos para fijar la posición. Sobre ellos va montado un regulador de altura fijo intercambiable, que también sirve como soporte para dotar de orientación al sensor. Este sistema de orientación se encarga de establecer el ángulo del sensor respecto del centro del

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

soporte, además de dotar de sujeción al mismo.

En el centro de la base (1), y en torno al cual gira todo el sistema, se encuentra un soporte giratorio portador del objeto a reconstruir, el cual es capaz de rotar el ángulo necesario y dar sujeción a los objetos.

Debido a que un gran número de sensores necesita calibrar su posición, se dispone de un soporte fijo en la base (1), a modo de panel de calibración, que permite incorporar diferentes patrones de calibración fácilmente intercambiables debido a su construcción.

Asimismo, se dispone de una unidad de control en la base (1) que se encarga de traducir los movimientos enviados por una tablet (4), de forma simplificada a cada uno de los sistemas de movimiento del soporte para alcanzar la posición deseada. Se busca una automatización total para lograr una gran precisión y repetitividad en el proceso.

Finalmente, la posibilidad de embarcar dos sensores (pasivos/activos) permite dotar al sistema BRAMUR3D de una gran flexibilidad y versatilidad en la toma de datos, ya que gracias al soporte se conocen los elementos de su posicionamiento espacial y angular, pudiendo fusionar ambos y así obtener un producto tridimensional híbrido (e.g. cámara RGB + cámara Infrarroja, cámara RGB + cámara térmica, dos cámaras RGB (estereoscopía), proyector de luz estructurada + cámara RGB, laser escáner + cámara RGB, etc,).

DESCRIPCIÓN

BRAMUR3D

Objeto de la invención

La presente invención se refiere al desarrollo de un brazo que permite resolver la orientación angular y espacial de uno o dos sensores con respecto al centro de un soporte, que además integra un sistema para girar el soporte lo que permite reconstruir tridimensionalmente y de forma completa cualquier objeto.

Antecedentes de la invención

La salida al mercado de sensores de bajo coste con diferentes propósitos (videojuegos, visión computacional, etc.) ha sido incesante en los últimos años, más si cabe con la reciente incorporación de sistemas de impresión 3D que se precisan como dato de entrada de modelos tridimensionales en formato CAD. La configuración de los diferentes sensores respecto al objeto a reconstruir, posiciones relativas entre los sensores y el objeto, es primordial en los sensores de bajo coste de cara a la calidad, precisión y fiabilidad de los resultados en el modelo generado. Ante la necesidad de poder probar varias combinaciones de la configuración y contar con la posibilidad de repetir la óptima para cada sensor y objeto se crea la necesidad de tener un soporte automático. En la actualidad existen diferentes soportes para estos sensores, los más utilizados son trípodes móviles. Estos trípodes permiten la orientación de la cámara, y aunque existen algunas soluciones automáticas, no son capaces de reproducir la misma posición con respecto a un objeto. Otros soportes son fijos, tienen una posición fija con respecto a un sensor, lo que limita la posible configuración del sensor para diferentes casos. En muchos casos también limita la utilización a un único sensor para el que está diseñado.

Con respecto al sistema central de mesa giratoria ya existen muchas soluciones para resolver ese problema, pero ninguno es

VNiVERSiDAD D SALAMANCA

OFICINA DE TRANSFERENCIA DE RESULTADOS DE INVESTIGACIÓN (OTRI)

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

capaz de transmitir la posición exacta de la mesa a cada uno de los sensores.

La mayor ventaja que proporciona el sistema BRAMUR3D es la automatización del proceso de configuración espacial de los sensores, la posibilidad de añadir varios sensores con diferentes configuraciones al sistema, el contar con movimiento relativo entre sensores y objeto actuando sobre cualquiera de ellos, y el conocimiento de la posición y orientación relativa entre los sensores y el objeto para todas las capturas realizadas. Como la mayoría de los sensores que este sistema puede incorporar, suelen tener una limitación en su campo de vista, abarcando ángulos entre 30º a 60º, este dispositivo también incorpora la capacidad de rotar el objeto a reconstruir alrededor de su vertical.

A continuación se relacionan varios documentos pertenecientes al estado de la técnica:

Documento	Referencia	Fecha
D01	SCANNER LIGHT STRUCTURE (http://www.multistation.com/SCANNER-LIGHT-STRUCTURE)	2013
D02	Photosimile 5000 - Automated 3D Product Photography Studio (http://ortery.com/Photography-Equipment/3D-Photography-Equipment/Photosimile-5000-Automated-3D-Product-Photography)	2010
D03	3D PhotoArm (http://ortery.com/Photography-Equipment/3D-Photography-Equipment/3D-PhotoArm-2000-3D-Product-Image-Equipment)	2010
D04	MarkerBot Digitalizer (http://store.makerbot.com/digitizer.html)	2013
D05	Solo Shot(http://soloshot.com/)	2012
D06	Virtucube (http://www.thingiverse.com/thing:52850)	2013

Descripción de la invención

A continuación se pasa a describir la invención BRAMUR3D que detalla cada uno de sus elementos.

1. Soporte central (Figura 2)

   En esta parte se encuentra la base (11) en la que se ubica la unidad de control (12), que se encarga de controlar todas las partes del soporte y la comunicación con una tablet exterior vía wifi. Asimismo, cuenta con un eje vertical en torno al que giran los brazos (13) y el soporte porta patrones de calibración (16).

   Sobre este eje se ubica la mesa rotatoria (14), que se encarga de girar el objeto a reconstruir, permitiendo incluso un giro completo de 360º, y contando con un pequeño dispositivo (15) de amarre para evitar los movimientos del objeto a escanear en la rotación.

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

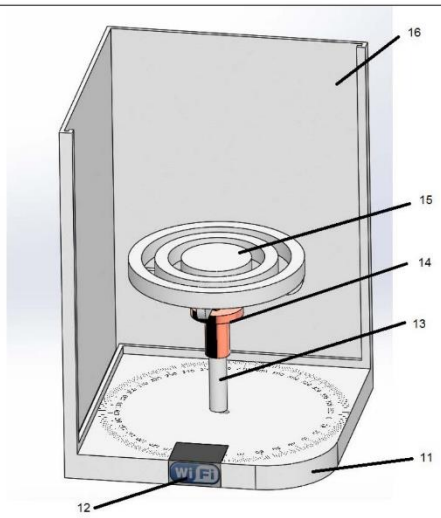


Figura 2. Soporte central base en detalle.

2. Brazos rotatorios (Figura 3)

En la base existe un eje vertical sobre el que se coloca la mesa rotatoria con los objetos, alrededor de este eje giran los brazos. Estos brazos se encargan de controlar el alejamiento de los sensores respecto al objeto y el ángulo que forman entre ellos y con respecto al objeto. Como se observa en la Figura 3, aunque el cálculo del ángulo es automático, se dispone de un visor para comprobarlo (23), ocurriendo lo mismo con la distancia o alejamiento al centro del eje (22). Una de las características más sobresalientes del brazo es su movilidad. Para poder comunicarse con los diferentes módulos que lo forman, añade unos railes (21) en los cuales se implementa la alimentación y señales de mando para los objetos. El sistema de comunicación consiste en una transmisión serie a través dela cual la unidad de control emite los mensajes para cada uno de los módulos. Lo módulos después interpretan la posición que tienen que adoptar.

VNiVERSiDAD
Đ SALAMANCA

OFICINA DE TRANSFERENCIA DE RESULTADOS DE INVESTIGACIÓN (OTRI)

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

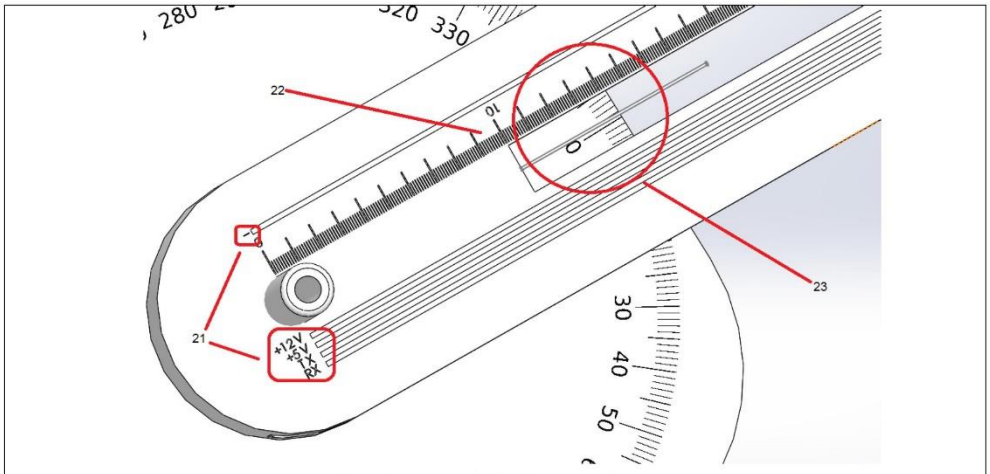


Figura 3. Detalle de los brazos rotatorios.

En la Figura 4 se puede apreciar de forma detallada el lateral de los brazos. El diseño de los mismos está configurado de forma que no tengan problemas de choque en el giro, ya que uno de ellos abraza al otro. También se puede obsevar (Figura 4) el engranaje de cremallera de que dispone cada brazo, que además de mejorar la sujeción, al ser como un raíl, permite desplazar el soporte horizontalmente según el alejamiento deseado.

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

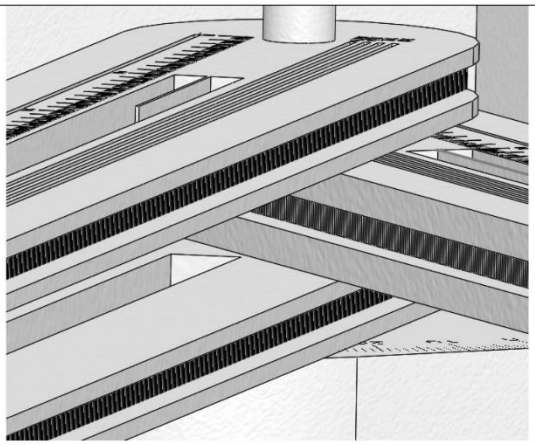


Figura 4. Detalle del engranaje lateral de los brazos.

La regulación del ángulo entre los sensores se produce a través de una pequeña rueda en el extremo de la misma (Figura 5). El ángulo se regula con el cálculo del arco de la circunferencia. Una controladora se encarga de traducir los mensajes que recibe de la unidad de control en movimiento. La precisión en el ángulo depende del tamaño del brazo usado y de la resolución del motor que mueve la rueda.

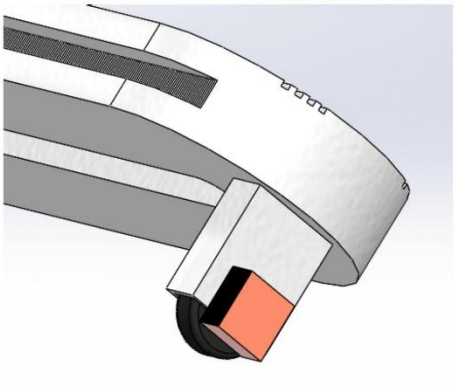


Figura 5. Detalle de la rueda de rotación.

VNiVERSiDAD D SALAMANCA

OFICINA DE TRANSFERENCIA DE RESULTADOS DE INVESTIGACIÓN (OTRI)

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

3. Soporte de los sensores.

Este soporte (Figura 6) se encarga de sujetar el sensor y posicionarlo en altura y ángulo respecto al objeto.

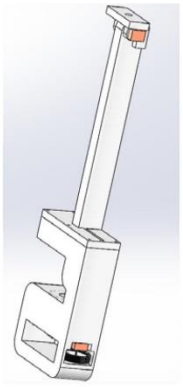


Figura 6. Soporte del sensor.

Entrando más en detalle (Figura 7), en la parte inferior del mismo se sitúa una controladora (31) encargada de mover el soporte a lo largo del brazo, dotando a este del alejamiento deseado de forma automática. Las ruedas metálicas (32) son las encargadas de comunicar los raíles del brazo con el soporte, proporcionándole energía y comunicación con la unidad de control.

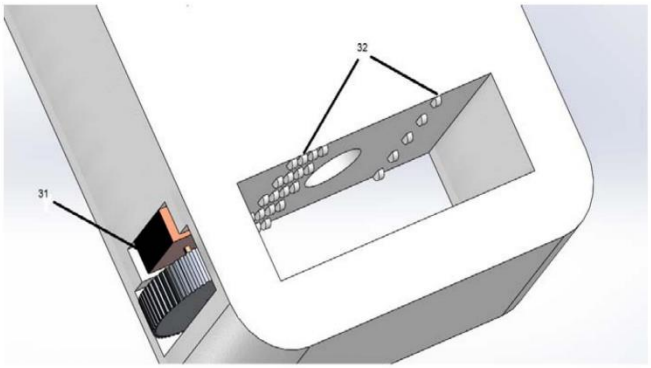


Figura 7. Detalle de la parte inferior del soporte sensor.

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

En la parte central del soporte (Figura 8) se encuentra la pieza intercambiable (33) encargada de regular la altura. Según el sensor y el objeto a reconstruir, se puede optar entre varias medidas de este soporte. Este soporte también lleva un pasador que engrana con el brazo para conseguir mayor estabilidad (34).

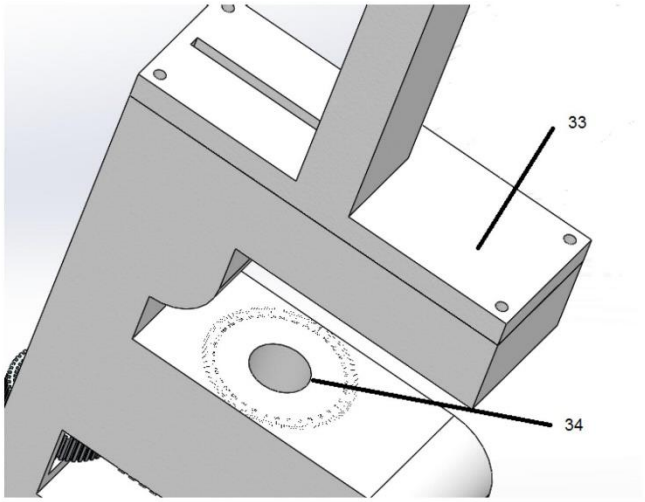


Figura 8. Detalle de la parte media del soporte sensor.

En la parte superior del mismo (Figura 9) se encuentra la sujeción del sensor (35), de posición fija para permitir la repetición de la configuración. Esta sujeción tiene la capacidad de regular el ángulo de cabeceo del sensor, para lo que utiliza una controladora con un motor (36). Este ángulo generado automáticamente por la unidad de control se puede contrastar de forma visual con el medidor de ángulos analógico que incorpora (37). Para la puesta a cero del mismo se utiliza un fin de carrera (38) que permite recalibrar la controladora.

VNiVERSiDAD D SALAMANCA

OFICINA DE TRANSFERENCIA DE RESULTADOS DE INVESTIGACIÓN (OTRI)

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

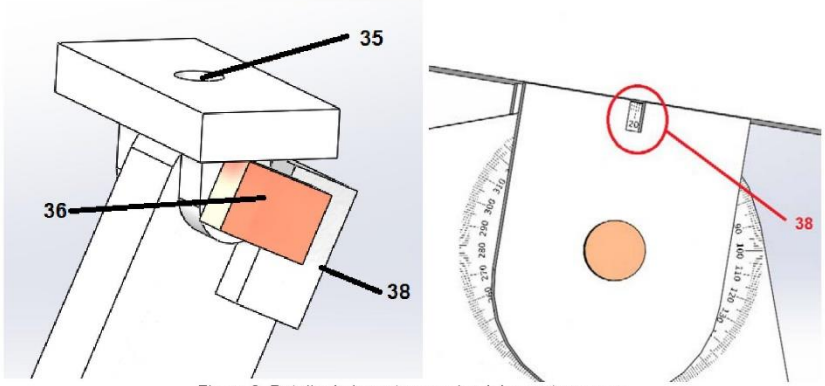


Figura 9. Detalle de la parte superior del soporte sensor.

4. Tablet controladora.

Para facilitar el control del dispositivo, la facilidad de configurarlo y la repetición de la configuración, el soporte es controlado a través de una tablet. Esta tablet cuenta con un software que permite al usuario gestionar el sistema, introduciendo el valor de cada grado de libertad para el soporte que a continuación se envía a la unidad de control vía wifi, lo que se traduce en dotar al sistema de una configuración automática. Esta tablet tendrá la capacidad de hacer persistir toda la información a través de un sistema gestor de proyectos en bases de datos y generar plantillas en base al aprendizaje del usuario adquirido para un tipo particular de sensores y tipos de objetos.

REIVINDICACIONES

1. Brazo para la reconstrucción 3D automática de objetos de reducido tamaño a partir de dos sensores iguales o diferentes que comprende:
   - Un soporte central (1) en el que se encuentra la unidad de control y que se encarga de controlar todos todas la partes del soporte. Asimismo, cuenta con una plataforma rotatoria en la que se disponen los objetos a reconstruir junto con un porta patrones de calibración.
   - Dos brazos articulados (2) encargados de controlar las posiciones relativas, distancia y orientación, entre los mismos y de estos con respecto al objeto.
   - Dos soportes de los sensores (3), encargados de sujetarlos y posicionarlos en altura y ángulo respecto al objeto.
   - Una tablet controladora para facilitar el control y configuración del dispositivo.
2. Brazo según reivindicación 1 caracterizado porque permite automatizar la posición espacial (alejamiento) y angular de los sensores con respecto al objeto a reconstruir en función de las restricciones o condiciones impuestas por el usuario.
3. Brazo según reivindicación 1 caracterizado por incorporar una plataforma rotatoria automatizada que controla y registra el ángulo de rotación del objeto en el mismo sistema de coordenadas que los sensores embarcados, permitiendo reconstruir cualquier parte del objeto.
4. Brazo según reivindicación 1 caracterizado por permitir embarcar diferentes tipos de sensores tanto pasivos (cámara) como activos (láser).

VNiVERSiDAD D SALAMANCA

OFICINA DE TRANSFERENCIA DE RESULTADOS DE INVESTIGACIÓN (OTRI)

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

5. Brazo según reivindicación 1 caracterizado por incorporar un sistema de control que permite enviar señales de disparo a los sensores para capturas múltiples de forma automática.
6. Brazo según reivindicación 1 caracterizado por incorporar dos soportes de sensor que permite posicionarlos en altura y ángulo con respecto al objeto.

Palabras Clave para efectuar búsquedas en bases de datos (en español y en inglés):

Español: brazo automatizado multifunción, multisensor, reconstrucción 3D, ingeniería inversa.

Inglés: multifunction automatic arm, multisensor, 3D reconstruction, reverse engineering.

6.- APLICACIONES DE LA INVENCIÓN

Describir brevemente las aplicaciones industriales de la invención e indique claramente problema que resuelve el dispositivo o procedimiento de dicha invención (máximo, 200 palabras)

La posibilidad de conocer con precisión la posición (espacial y angular) de diferentes sensores respecto a un objeto en el mismo sistema de coordenadas permite abordar la reconstrucción automática 3D de objetos usando diferentes sensores pasivos o activos (i.e. cámara fotográfica, luz estructurada, láser escánaer, cámara térmica, cámara multiespectral y combinación de varios), y además permite integrar información procedente de ambos sensores a los modelos 3D resultantes, para, por ejemplo, dotar de textura a los modelos a partir de las imágenes de alta resolución capturadas. Asimismo, la posibilidad del gran número de configuraciones del brazo permite realizar diferentes pruebas o calibraciones del mismo auxiliándose de los patrones de calibración que se pueden añadir, obteniendo la configuración óptima para la captura de datos. Conseguida la configuración óptima para cada sensor, este sistema permite repetir cada configuración el número de veces que sea necesario. Las aplicaciones de este brazo son numerosas, destacando:

- Generación de modelos 3D de precisión y con textura en formato CAD que sirvan como dato de entrada a impresoras 3D.
- Análisis dimensional y reconstrucción tridimensional de objetos.
- Prototipado y procesos de ingeniería inversa en industria.
- Calibración y configuración de sensores
- Hibridación de información sobre objetos 3D utilizando la información procedente de distintos sensores.
- Configuración automática de sensores.

Sectores a los que va dirigido (señalar tanto el sector que lo debería producir como el que lo utilizaría):

SECTOR	Producción	Utilización

VNiVERSiDAD D SALAMANCA

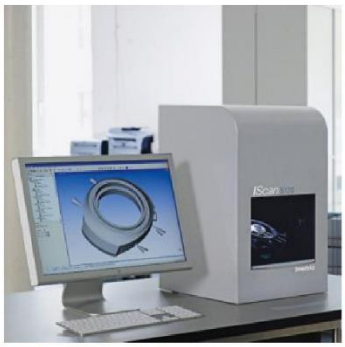
OFICINA DE TRANSFERENCIA DE RESULTADOS DE INVESTIGACIÓN (OTRI)

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

0. Agricultura	☐	☐
1. Industrias extractivas y del petróleo	☐	☐
2. Alimentación, bebidas, tabaco	☒	☐
3. Textil, confección, cuero y calzado	☒	☐
4. Madera y corcho	☒	☐
5. Papel, edición, artes gráficas y reproducción	☒	☐
6. Química y farmacia	☐	☐
7. Caucho y materias plásticas	☒	☐
8. Productos minerales no metálicos diversos	☐	☐
9. Metalurgia y fabricación de productos metálicos	☒	☐
10 Maquinaria y equipo mecánico	☒	☒
11.Material y equipo eléctrico, electrónico y óptico	☒	☐
12. Material de transporte	☐	☐
13.Industrias manufactureras diversas. Reciclaje	☒	☐
14.Energía y agua	☐	☐
15.Construcción	☒	☐
16.Comercio y hostelería	☐	☐
17.Transportes y comunicaciones	☐	☒
18.Inmobiliarias, alquileres y servicios a empresas	☐	☐
19.Servicios públicos, sociales y colectivos	☐	☐

Por favor, enumere los productos alternativos a su invención que ya existen en el mercado actualmente:

D01: Dispositivo de captura de información 3D a través de luz estructurada. Es capaz de introducir pequeños giros manualmente y está limitado a un único sensor. La posición de los sensores es única, moviendo solo el objeto.



D02: Dispositivo de captura de imágenes entorno a un objeto. Sólo es capaz de utilizar un tipo de sensor, cámaras fotográficas réflex. Es capaz de tomar fotografías desde una envolvente en torno al objeto. También es capaz de rotar el objeto. La problemática que presenta este sistema es que dispone de un único tipo de sensor y sólo incorpora cuatro grados de libertad. Carece de repetibilidad.

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

D03: Este dispositivo dispone de seis grados de libertad y la posibilidad de incorporar diferentes sensores. Es un dispositivo fijo o móvil. No es capaz de reproducir configuraciones entre disparos. El movimiento del sensor se realiza de forma manual. Carece de repetibilidad.

D04: Este dispositivo dispone de un único tipo de sensor (luz estructurada). Dispone de una mesa rotatoria entorno a la cuál gira el objeto. Permite una única configuración y un único sensor. La mayor ventaja de este sensor es su repetitibilidad.

VNiVERSiDAD D SALAMANCA

OFICINA DE TRANSFERENCIA DE RESULTADOS DE INVESTIGACIÓN (OTRI)

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

D05: Es un dispositivo capaz de moverse automáticamente siguiendo un objeto definible. Tiene la capacidad de embarcar diferentes sensores. No mantiene una posición fija con respecto al objeto y sólo se puede embarcar un sensor de cada vez. Carece de repetibilidad.

D06: Es un dispositivo diseñado para un escáner de luz estructura, con una única configuración. Dispone de dos soportes para diferentes sensores. Sólo dispone de un grado de libertad en la mesa rotatoria que se encarga de mover el objeto. Sólo puede embarcar un tipo de sensor y una única configuración. Carece de repetibilidad.

VNiVERSiDAD
Đ SALAMANCA

OFICINA DE TRANSFERENCIA DE RESULTADOS DE INVESTIGACIÓN (OTRI)

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

Indique qué ventajas técnicas tiene la invención respecto a los citados productos (máximo 100 palabras).

Las ventajas técnicas de la invención respecto a los productos reseñados en el estado de la técnica se centran en:

1) **Brazo Multisensor**, al poder emplearse tanto sensores pasivos como activos de diferentes tipos, principalmente cámaras digitales y escáneres.
2) **Automático**, el movimiento (angular y espacial) es producido automáticamente por el soporte.
3) **Amigable**, el sistema de control permite posicionar los sensores en base a los requerimientos del usuario. Resolución automática de movimientos.
4) **Metrológico**, es capaz de fijar posiciones en los objetos con precisiones por debajo del milímetro en distancia y en torno al minuto en ángulos, precisión similar por la pequeña distancia al objeto.
5) **Flexible**, al poder posicionar los sensores con todos los grados de libertad necesarios (acimutal, cenital, alejamiento, distancia base entre sensores, elevación) entre ellos y con respecto al objeto a modelizar.
6) **Repetibilidad**, ya que al estar automatizado permite repetir tomas en la misma posición (con precisión de minutos para el caso de los ángulos y de milímetros parea el caso de las distancias).
7) **Bajo coste** del soporte.
8) **Portabilidad y robustez**, dada su ligereza y diseño compacto.
9) **Gestión informatizada,** al contar con un software instalable en dispositivos hardware de bajo coste, tipo

VNiVERSiDAD D SALAMANCA

OFICINA DE TRANSFERENCIA DE RESULTADOS DE INVESTIGACIÓN (OTRI)

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

tablet, que incorpora la gestión de los trabajos en forma de proyecto, generación de plantillas de configuraciones mejorables en base a la experiencia del usuario con cada tipo de trabajo en base a los tipos de sensores y objetos, etc.

La invención se considera nueva porque:

- ☒ no se ha encontrado nada igual en bancos de datos de patentes
- ☒ no se ha encontrado nada igual en la bibliografía científica consultada
- ☒ no se ha encontrado nada igual en un informe de búsqueda completo

7.- GRADO DE DESARROLLO DE LA INVENCIÓN.

Elegir, entre estas opciones, la o las que más se aproximen al grado de desarrollo de la invención:

- ☐ se ha realizado en laboratorio, exclusivamente.
- ☐ se ha realizado ensayo en planta piloto.
- ☒ existe prototipo preparado para su desarrollo y comercialización.
- ☐ habría que realizar una serie de desarrollos para su comercialización o implantación industrial

En el caso de que sea necesario realizar su desarrollo para la explotación comercial, éste tendría:

dificultad técnica,

☐ elevada ☒ normal ☐ baja

coste económico,

☐ elevado ☐ medio ☒ bajo

8.- GRADO DE DIFUSIÓN DE LA INVENCIÓN

¿Se ha difundido previamente el objeto de la invención ?

☐ SI ☒ NO

En caso afirmativo, indicar medio de difusión, fecha y contenido (Tesis, publicaciones, congresos…)

9.- EXPLOTACIÓN Y COMERCIALIZACIÓN DE LA PATENTE

Se considera que sería un producto (procedimiento) con posible éxito comercial:

☐ elevado ☒ medio ☐ bajo

¿Se ha contactado con alguna empresa para su posible explotación?

☒ SI ☐ NO

VNiVERSiDAD D SALAMANCA

OFICINA DE TRANSFERENCIA DE RESULTADOS DE INVESTIGACIÓN (OTRI)

Casa del Bedel, C/ Cardenal Pal i Deniel, 22
37008 - Salamanca
Tel . (34) 923 29 44 90 Fax . (34) 923 29 46 64

En caso afirmativo, ¿con cuál? PROYESTEGAL; SITOP

En caso negativo, o si procediese, conteste la siguiente pregunta

¿Conoce alguna empresa que pudiera estar interesada ?

☒ SI ☐ NO

¿podría indicar cuál o cuáles? PROYESTEGAL; SITOP

El mercado de la patente es:

☒ exclusivamente nacional

☐ internacional (señalar):

☐ EE.UU. ☐ Europa ☐ Japón ☐ América del Norte
☐ Africa ☐ Australia ☐ Otros :

APÉNDICE C: Software

Título: GSMOD – Gaming Sensor Modeler

Número de inscripción: SA-137-13

Autores: Manuel Gesto Díaz, Diego Guerrero Sevilla, David Hernández López, Diego González Aguilera, Jesús Fernández Hernández, Pablo Rodríguez Gonzálvez

Propietario: Universidad de Salamanca

Fecha de inscripción: 30 de Julio de 2013

Descripción

La emergencia de los “gaming sensors” o sensores de videojuegos como dispositivos low-cost de modelado 3d representan una alternativa de interés general para múltiples aplicaciones de ingeniería y arquitectura en las que el grado de precisión no es excesivo y se requiere acometer procesos de ingeniería inversa que permitan el modelado 3d de objetos o escenarios de tamaño reducido. Uno de los mayores inconvenientes en el manejo de estos sistemas radica en la existencia de una herramienta software compacta y sencilla que permita bajo los sistemas operativos Windows y Linux acometer el proceso de captura y procesamiento de los datos para deparar un modelo 3d a escala.

La mayor ventaja del software es que permite trabajar con sistemas gaming sensor para poder utilizarlos en los procesos de escaneo y modelado 3D con un coste muy reducido y con resultados muy aceptables para muchas aplicaciones de ingeniería y arquitectura.

Sector al que va dirigido: cualquier empresa de ingeniería y arquitectura, en la que se requiera el trabajar con dispositivos de escaneo láser para modelizar pequeños objetos o escenarios.

Título: ReTrack

Número de inscripción: SA-47-15

Asiento registral: 00/2015/2739

Autores: Manuel Gesto Díaz, Diego González Aguilera, Pablo Rodríguez Gonzálvez

Propietario: Universidad de Salamanca

Fecha de inscripción: 26 de Febrero de 2015

Descripción

ReTrack (Recognition and Tracking System), es un sistema que consiste en el reconocimiento y seguimiento de objetos utilizando un gaming sensor. El sistema es alimentado con modelos CAD, de forma que a través de un gaming sensor y las imágenes de profundidad que proporciona el mismo se pueda llevar a cabo el reconocimiento de determinados objetos (existentes en la librería CAD) y su seguimiento en tiempo real.

Su originalidad reside en que garantiza automatismo y facilidad de uso proporcionando una utilidad no presente en otros softwares existentes.

Es de interés para empresas relacionadas con las HMI (Human Machine Interface), para empresas que proveen servicios a personas con ciertas discapacidades (visual principalmente) y empresas de automatización de procesos que necesitan el reconocimiento de objetos.