



**Universidad**  
**Zaragoza**

Trabajo Fin de Grado

# **Evaluación de ORBSLAM en Secuencias de Endoscopia Médica**

Autor

IÑIGO CIRAUQUI VILORIA

Director

JOSE MARÍA MARTÍNEZ MONTIEL

Departamento de Informática e Ingeniería de Sistemas

Escuela de Ingeniería y Arquitectura

2016



## DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD

(Este documento debe acompañar al Trabajo Fin de Grado (TFG)/Trabajo Fin de Máster (TFM) cuando sea depositado para su evaluación).

D./D<sup>a</sup>. Iñigo Cirauqui Viloría

con nº de DNI 73119875 T en aplicación de lo dispuesto en el art.

14 (Derechos de autor) del Acuerdo de 11 de septiembre de 2014, del Consejo

de Gobierno, por el que se aprueba el Reglamento de los TFG y TFM de la

Universidad de Zaragoza,

Declaro que el presente Trabajo de Fin de (Grado/Máster)

Grado en Ingeniería de Tecnologías Industriales, (Título del Trabajo)

Evaluación de ORBSLAM en Secuencias de Endoscopia Médica

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

es de mi autoría y es original, no habiéndose utilizado fuente sin ser citada debidamente.

Zaragoza, 22 de abril de 2016

Fdo: \_\_\_\_\_

# Evaluación de ORBSLAM en Secuencias de Endoscopia Médica

## RESUMEN

Se presenta la evaluación de un sistema de Visual SLAM, el ORBSLAM, en escenas in vivo de endoscopia en animales. Un sistema de SLAM visual procesa la secuencia de imágenes que toma una cámara que se mueve por un entorno desconocido siguiendo una trayectoria también desconocida. A partir del procesamiento de la secuencia de imágenes el sistema proporciona, en tiempo real, tanto un mapa 3D de la escena como la posición de la cámara respecto de este mapa.

El ORBSLAM es un sistema diseñado para escenas de robótica móvil donde predominan los elementos rígidos. Se detalla la resintonía del sistema para adaptarlo al tratamiento de escenas que contienen elementos no rígidos. También se aborda la calibración del sistema.

El código original del sistema se modifica para visualizar y recoger datos de forma detallada en los dos pasos que están limitando el funcionamiento del sistema en escenas de endoscopia: el tracking de la cámara y gestión de puntos del mapa.

Posteriormente se recopilan datos estadísticos que definen el funcionamiento en 19 secuencias tomadas in-vivo durante operaciones de endoscopia en animales.

Tras un estudio se detallan los condicionantes para un funcionamiento correcto y se proponen diversas líneas de trabajo futuro con las que continuar la investigación, en vistas de acabar logrando un sistema SLAM visual específico para secuencias médicas.

# Índice

<b>1. Introducción</b>	<b>6</b>
1.1. Motivación . . . . .	6
<b>2. Metodología</b>	<b>8</b>
<b>3. Descripción de ORBSLAM</b>	<b>9</b>
3.1. Inicialización . . . . .	10
3.2. Emparejamiento en el tracking . . . . .	11
3.3. Gestión del mapa . . . . .	12
<b>4. Puesta en marcha</b>	<b>13</b>
4.1. Calibración . . . . .	13
4.2. Sintonía . . . . .	17
<b>5. Experimentos</b>	<b>19</b>
5.1. Cavidad abdominal: hígado y bazo . . . . .	21
5.1.1. Tracking . . . . .	21
5.1.2. Gestión del mapa . . . . .	22
5.2. Cavidad abdominal: hígado y vesícula biliar . . . . .	23
5.3. Cavidad abdominal: intestino . . . . .	23
5.4. Endoscopia digestiva alta: esófago y estómago . . . . .	26
5.5. Bucles en la cavidad abdominal . . . . .	28
<b>6. Resultados</b>	<b>30</b>
6.1. Discusión . . . . .	30
6.2. Conclusiones . . . . .	31
6.3. Trabajo futuro . . . . .	32
<b>7. Referencias</b>	<b>33</b>
<b>Anexos</b>	<b>35</b>
<b>A. Leyenda de las ventanas</b>	<b>35</b>
A.1. Ventana 1: Tracking . . . . .	35
A.2. Ventana 2: Gestión del mapa (edad de los puntos) . . . . .	35
A.3. Ventana 3: Gestión del mapa (borrado de puntos) . . . . .	35

<b>B. Secuencias del estudio</b>	<b>37</b>
B.1. Endoscopio 1 . . . . .	37
B.2. Endoscopio 2 . . . . .	37
B.3. Endoscopio 3 . . . . .	38
B.4. Endoscopio 4 . . . . .	38
<b>C. Datos experimentales</b>	<b>40</b>
C.1. Cavidad abdominal: hígado y bazo . . . . .	40
C.2. Cavidad abdominal: hígado y vesícula biliar . . . . .	40
C.3. Cavidad abdominal: intestino . . . . .	40
C.4. Endoscopia digestiva alta: esófago y estómago . . . . .	40
C.5. Bucles en la cavidad abdominal . . . . .	40

# 1. Introducción

Se denomina VSLAM, *Visual Simultaneous Localization And Mapping*, a un conjunto de algoritmos empleados principalmente en robótica y realidad aumentada para generar un mapa tridimensional del entorno a partir de imágenes tomadas mediante cámaras en movimiento que observan la escena. Al mismo tiempo que se construye el mapa, se proporciona una estimación de la localización de las cámaras respecto de este mapa. Estos sistemas permiten que genere un mapa de un entorno desconocido a la vez que se localiza el sensor, siendo uno de los elementos básicos para la interacción de un robot con su entorno. En este trabajo nos focalizamos en los sistema de SLAM visual que emplean una única cámara monocular.

Davison [1] presenta la primera demostración de un sistema de VSLAM monocular funcionando en tiempo real en escenas genéricas. El sistema estaba basado en un Filtro de Kalman Extendido (Extended Kalman Filter, EKF) [2] que llevaba a cabo una actualización probabilista del mapa generado. El sistema tenía una inicialización robusta, pero estaba limitado en cuanto al tamaño del mapa que podía gestionar en tiempo real, compuesto solo por unos pocos cientos de puntos. Avances sobre el sistema EKF SLAM, han sido presentados por [3, 4], aumentando la robustez y la capacidad de operar en exteriores.

Un salto significativo en los sistemas de VSLAM monocular llega con PTAM (*Parallel Tracking and Mapping*) [5], un método basado en keyframes. El algoritmo propuesto era capaz, en tiempo real, de realizar todas las etapas del proceso de reconstrucción fotogramétrica de la escena [6]: emparejamiento de puntos, orientación de la cámara, inicialización y optimización no lineal (*Bundle Adjustment*, BA) del mapa. El sistema puede manejar en tiempo real mapas de miles de puntos y localizar la cámara de forma muy precisa.

Recientemente el sistema ORB-SLAM [7], mejora las prestaciones de PTAM al añadirle cierres de bucle y una relocalización robusta en tiempo real. Es relevante que los puntos empleados para construir el mapa se basan en puntos de imagen ORB [8], que son invariantes a la rotación y a cambios de escala, con un rendimiento similar a los puntos SIFT [9] –método de referencia en detección de puntos de interés– pero a menor coste computacional.

## 1.1. Motivación

Los métodos de VSLAM se emplean con escenas tomadas desde robots móviles tanto en entornos interiores como exteriores, buscando crear un mapa de un entorno desconocido. La endoscopia es una técnica diagnóstica que consiste en la introducción

de una única cámara (endoscopio) a través de un orificio natural o provocado para visualizar una cavidad corporal y recopilar información sobre esta.

Existe una analogía entre ambas situaciones, en las que una única cámara recorre una trayectoria en un entorno del que no se tiene información anterior. A raíz de esto surge la idea de emplear sistemas VSLAM en escenas de endoscopia médica. El EKF-VSLAM ha sido empleado satisfactoriamente con imágenes médicas de cirugía abdominal [10] proporcionando mediciones relevantes para la intervención, o implementando funciones de realidad aumentada como son las anotaciones en tiempo real [11].

Nuestra contribución es la aplicación del sistema ORB-SLAM, diseñado para escenas rígidas propias de la robótica, en secuencias de endoscopia médica que que presenten un comportamiento no rígido. Nuestra metodología consiste en hacer la suposición de que la escena es rígida y analizar cuales son los límites de esta aproximación. Nos focalizamos en el análisis de qué puntos se incluyen en el mapa, y en el emparejamiento de estos puntos a lo largo de la secuencia. También prestaremos atención a qué puntos emplea la cámara para localizarse respecto del mapa.

Nuestro objetivo es identificar cuales son los factores clave que limitan el funcionamiento del ORBSLAM en este tipo de escenas. Nuestras conclusiones están avaladas por una verificación experimental sobre secuencias de endoscopia in vivo del abdomen de un cerdo. Se establecen como objetivos intermedios:

1. Adaptación y sintonía del código para el procesamiento de las secuencias de prueba que incluyen escenas con elementos no rígidos.
2. Análisis del comportamiento del sistema, identificando los factores limitantes en la búsqueda de emparejamientos y en la creación de puntos del mapa.
3. Buscamos analizar especialmente el comportamiento de las áreas no rígidas de la escena, en vistas a un trabajo futuro de segmentación y tratamiento de áreas no rígidas.

Hasta donde nosotros sabemos es la primera vez que se prueba el sistema ORBSLAM en secuencias de endoscopia médica. En la verificación experimental analizamos un total de 19 secuencias de diferentes tipos de endoscopia médica para dar respaldo experimental a las conclusiones.

La sección 2 se presenta la metodología de trabajo. La sección 3 se describe el sistema ORBSLAM, resaltando las etapas del proceso relevantes en nuestro análisis. A continuación, en la sección 4 se describen los ajustes que permiten mejorar el funcionamiento para escenas no rígidas de endoscopia. La sección 5 presenta el análisis

experimental, en el que se describen los resultados obtenidos al probar el sistema con las 19 secuencias de endoscopia médica. Finalmente la sección 6 está dedicada a las conclusiones y el trabajo futuro.

Por último se presentan 3 anexos con la leyenda de la información mostrada por el software, un listado de las secuencias empleadas y su descripción, y para concluir, se completa el estudio experimental con datos adicionales.

## 2. Metodología

Con el fin de lograr los objetivos mencionados, el trabajo a realizar se estructura como sigue:

1. Se dispone de un conjunto de secuencias in-vivo tomadas en el interior de la cavidad abdominal de un cerdo. Las secuencias tienen diferentes niveles de dificultad.
2. Se dispone de una versión del ORBSLAM sin modificar [github.com/raulmur/ORB\\_SLAM](https://github.com/raulmur/ORB_SLAM).
3. Configuración y puesta en marcha de un entorno de desarrollo C++ sobre Linux, OpenCV, ImageMagic, ROS (Robot Operating System) y ORBSLAM.
4. Adaptación de las secuencias de imágenes reduciendo la resolución.
5. Calibración del endoscopio según el modelo de cámara de 4 parámetros de distorsión de ORBSLAM.
6. Sintonía del ORBSLAM para el procesamiento de las secuencias.
7. Análisis experimental del proceso de emparejamiento en el tracking. Se modifica el código para visualizar y cuantificar el efecto cada una de las etapas del proceso de emparejamiento.
8. Análisis experimental del proceso de creación de nuevos puntos en el mapa. Se modifica el código para visualizar y cuantificar el efecto cada una de las etapas del proceso de emparejamiento.
9. Análisis de resultados y conclusiones.
10. El código se hace público con licencia GPLv3.0



### 3. Descripción de ORBSLAM

ORBSLAM [7] es un sistema de Visual SLAM que basa su funcionamiento en la detección y emparejamiento de puntos imagen ORB. El sistema lleva a cabo varias tareas diferenciadas:

tracking.- Se asume que se dispone de una mapa. La cámara toma una imagen, sobre la imagen se detectan los puntos que ya están incluidos en el mapa. A partir de estos puntos se estima la posición de la cámara respecto del mapa.

mapping.- Proceso que construye el mapa de la escena a partir de las imágenes tomadas hasta este momento. Se necesita identificar puntos correspondientes entre un conjunto de imágenes seleccionadas. Los puntos emparejados son los puntos del mapa, de los que tenemos información 3D. Las imágenes seleccionadas las denominaremos keyframes. El proceso que estima y refina la posición de los puntos y de los keyframes es una optimización no lineal llamada Bundle Adjustment (BA).

relocation.- En la operación normal, para estimar la posición de la cámara se dispone de la posición que esta misma cámara tenía en el frame anterior, hace unos 30 ms, lo que simplifica el proceso. Puede darse el caso de que esta posición anterior no este disponible y haya que hacer una localización desde cero, este proceso se llama relocation. También se conoce como el caso del robot secuestrado.

gestión de mapa.- El mapa crece conforme la cámara explora zonas nuevas. Hay que decidir cuando se añaden keyframes al mapa, y cuando se añaden nuevos puntos que mapean las nuevas áreas exploradas. La gestión del mapa también incluye la eliminación de puntos y de keyframes.

cerrado de bucle.- A medida que vamos explorando nuevas zonas el mapa va creciendo. Si volvemos a una zona explorada previamente el sistema debería detectarlo. Es proceso es el cerrado de bucles.

inicialización.- Cuando se arranca el sistema no disponemos de mapa. La inicialización es la creación de la primera parte del mapa a partir los dos primeros keyframes que han de ser seleccionados y emparejados. Después se realiza la primera estimación de los puntos 3D del mapa.

Los puntos de interés en las imágenes, o keypoints, son la espina dorsal del funcionamiento de cada una de las etapas del sistema. Son localizados mediante el detector oFAST multiescala [8], y para cada uno de ellos se dispone de un descriptor binario



cializa con homografía si este supera un umbral que ha demostrado empíricamente su capacidad de distinguir escenas planas o no:

$$Ratio_H = \frac{S_H}{S_H + S_F} > 0,45 \quad (1)$$

Con el modelo escogido se crea un mapa inicial. En este trabajo, por el tipo de escenas que manejamos, hemos forzado que se asuma que la escena nunca es plana, por lo que no se emplea el modelo basado en homografía. Gracias a ello, el sistema arranca con mayor rapidez que si se consideraran ambos casos.

### 3.2. Emparejamiento en el tracking

El tracking se encarga, para cada frame, de localizar la cámara respecto al mapa disponible; para ello se detectan los puntos de interés FAST sobre la imagen, se obtienen sus descriptores ORB, y se buscan emparejamientos entre estos y los puntos del mapa:

1. Proyección del mapa sobre el frame actual: se descartan los puntos cuya proyección se encuentre fuera de los límites de la imagen, y aquellos que por la orientación de su normal serían vistos con demasiado escorzo desde la posición actual de la cámara.
2. Cálculo de la distancia del punto al centro óptico de la cámara, con esta se comprueba que el punto se encuentra en la escala correcta (región invariante a la escala), y si se ha producido un cambio respecto de la escala en la que se inicializó el punto.
3. Todos los puntos que cumplen lo anterior se definen como puntos del mapa predichos en el frame, y se consideran candidatos para emparejarse en la imagen. Para cada uno de ellos disponemos de una hipótesis de dónde se debería ver en el frame actual y con qué descriptor ORB.
4. Búsqueda de emparejamientos: se compara el ORB del punto del mapa candidato con los ORB de puntos detectados sobre la imagen, considerando sólo puntos próximos al punto del mapa reproyectado y en el mismo nivel de escala. Si hay más de uno en la vecindad, se selecciona el más similar.
5. Ajuste de la posición de la cámara mediante una optimización no lineal de los puntos emparejados. Calculando cuando el residuo del error de reproyección después de optimizar se encuentra por encima de un umbral, se identifican aquellos emparejamientos que no se corresponden con una escena rígida.

En la Figura 2 se muestra una representación sintetizada del proceso de tracking.

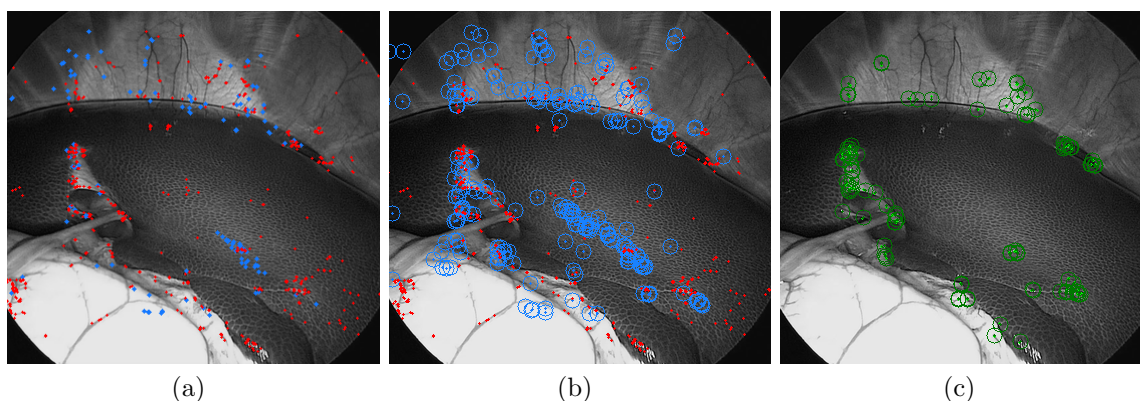


Figura 2: Proceso de tracking: (a) se detectan ORBs en el frame actual ( $\bullet$ ), y se proyecta el mapa sobre la imagen ( $\bullet$ ). (b) Se buscan emparejamientos para cada punto del mapa con puntos de la imagen que se encuentren en una región próxima a este y en el mismo nivel de escala, región definida en la figura por el círculo vacío, ( $\circ$ ). (c) Se selecciona el par mappoint-keypoint mas similar y se empareja ( $\odot$ ). Posteriormente se ajustará la posición de la cámara con estos emparejamientos.

### 3.3. Gestión del mapa

En ORBSLAM la gestión del mapa esta basada en keyframes, frames seleccionados cuya combinación proporciona suficiente información para construir un mapa, sin necesidad de frames intermedios. Si el frame actual muestra un punto de vista diferente a los otros keyframes ya incluidos en el mapa (un diferente conjunto de puntos del mapa presentes en la escena vista, Figura 3), este frame se convierte en keyframe y se emplea para crear nuevos puntos del mapa:

1. Se compara el keyframe recién creado con otros keyframes que también muestran puntos de la escena vista. Se emplea la geometría epipolar para seleccionar puntos susceptibles de ser emparejados.
2. Se comparan descriptores ORB, se emparejan, y se triangula el punto emparejado, comprobando que se encuentre delante del plano de la imagen en ambas cámaras, la existencia de suficiente paralaje, bajo error de reproyección, y consistencia de escala con los puntos de origen. Figura 3.
3. El Bundle Adjustment optimiza la posición de todos los keyframes y de todos los puntos del mapa que han participado en las etapas anteriores. Mediante un análisis de residuos se descartan los puntos que no concuerdan con una escena rígida.
4. Los nuevos puntos del mapa se evalúan durante 3 keyframes desde su creación, una dura prueba de estrés comprueba si el tracking es capaz de emparejarlos, y

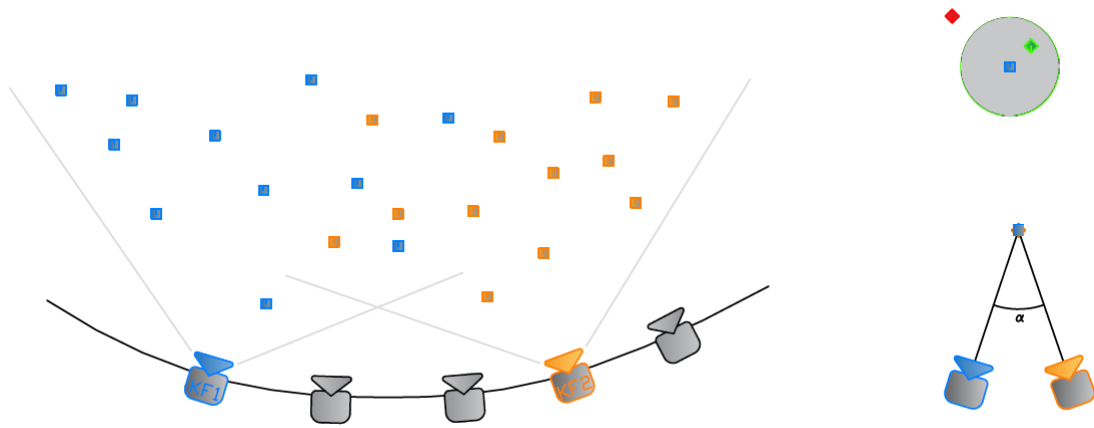


Figura 3: (izquierda): Frames a lo largo de una trayectoria: de todos ellos solo los dos coloreados se convierten en keyframes pues estos definen un cambio de la escena: el centro de la imagen es observable desde los dos keyframes, pero los extremos solo desde una de ellas. (superior derecha) Error de reproyección: para un punto del mapa proyectado en la imagen (azul) solo los puntos contenidos dentro del área de búsqueda (verde) pueden ser emparejados. (inferior derecha) Paralaje: el ángulo formado por los rayos que unen un punto en 3D con los centros ópticos de las cámaras que lo definen ha de ser mayor que un mínimo

si son observables en otros keyframes.

- Tras 3 keyframes los puntos supervivientes se vuelve permanentes y solo podrán ser borrados si en algún momento se observan desde menos de 3 keyframes, lo que puede ocurrir si algún keyframe se elimina; o si el ajuste robusto detecta repetidamente que el comportamiento del punto no se corresponde con el de una escena rígida.

La política del sistema se basa en ser generosos en cuanto a la creación de puntos del mapa, pero muy estrictos en su evaluación inicial, asegurando que solo puntos de alta calidad permanezcan en el mapa.

La evaluación que la función de ajuste robusto realiza con los emparejamientos en el tracking y con los puntos del mapa que se esta intentando crear, hace que una de las características principales de ORBSLAM sea una reducida tasa de puntos espurios.

## 4. Puesta en marcha

### 4.1. Calibración

Para un funcionamiento adecuado, ORBSLAM requiere un input compuesto por los parámetros intrínsecos de la cámara, que según (2) permiten conocer la posición de

puntos del entorno sobre el plano de la imagen, y por los coeficientes que definen la distorsión que causa la lente (3).

$$\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (2)$$

$$\text{Coeficientes de distorsión} = [ k_1 \quad k_2 \quad p_1 \quad p_2 \quad k_3 ] \quad (3)$$

El sistema esta configurado para funcionar con dos coeficientes de distorsión radial y los dos de distorsión tangencial, se ha encontrado que los segundos son dos órdenes de magnitud menor que los primeros para las cámaras empleadas, por ello se decide emplear solo los dos primeros coeficientes de distorsión radial, llevando a cabo una optimización que fije el resto en 0, haciendo que la posición final de un punto en la imagen se calcule según (4).

$$\begin{bmatrix} x_{\text{corregido}} \\ y_{\text{corregido}} \end{bmatrix} = (1 + k_1 r^2 + k_2 r^4) \begin{bmatrix} x \\ y \end{bmatrix} \quad (4)$$

Es necesario obtener los parámetros de calibración para 4 endoscopios diferentes, identificados en la Tabla 1 para facilitar referirnos a ellos en el estudio experimental. Se dispone de grabaciones en alta resolución (1920x1080) de patrones de calibración que han de ser procesados:

1. Extracción de los frames de las secuencias de calibración y selección de aquellos más adecuados para extraer los parámetros. Los vídeos no muestran posiciones de cámara ideales para la calibración, con lo que nos limitamos a escoger aquellos en los que se tenga un cambio sustancial en cuanto al punto de vista del patrón como son los mostrados en la Figura 4.
2. Reducción de la resolución de las imágenes a aquellas con las que se trabajara en ORBSLAM, 960x540 y 640x360 píxeles.
3. Desarrollo de código en base C++ implementando funciones de OpenCV para llevar a cabo la calibración, señalando explícitamente que el programa retorne solo dos coeficientes de distorsión radial, fijando en cero  $k_3$ ,  $p_1$  y  $p_2$ . El código se encuentra disponible en <http://www.github.com/cirauqui21/calibration>.
4. Procesado de las imágenes para obtener los parámetros mostrados en las Tablas 2 y 3.
5. Se verifica que la calibración es adecuada comparando los parámetros obtenidos con los proporcionados por IRCAD France (Instituto de Investigaciones contra el

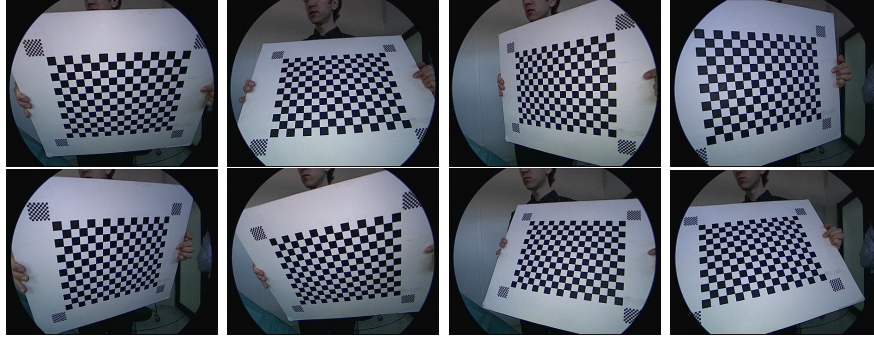


Figura 4: Frames seleccionados para obtener la calibración

Cáncer del Aparato Digestivo, Francia), de donde proceden las secuencias. En la Tabla 2 se presenta una comparación entre sus valores y los nuestros, apreciando mínima diferencia.

Los parámetros obtenidos solo son válidos para la configuración del endoscopio en que es grabado el patrón de calibración, siendo invalidados si durante la operación se alteran el zoom o el enfoque del endoscopio. En las secuencias disponibles no podemos asegurar que esto no haya ocurrido antes de tomar las secuencias de endoscopia, por ello se propone llevar a cabo un ajuste manual de dichos parámetros, procedimiento que también permite obtener una calibración válida para secuencias en las que no se disponga de patrones de calibración:

1. Si se dispone de vídeo de calibración se emplean los parámetros obtenidos anteriormente como punto de partida; si no se dispone de vídeo de calibración se supone un comportamiento semejante entre endoscopios y se establecen como valores iniciales de calibración los obtenidos para el endoscopio 1, los cuales se optimizarán para el endoscopio tratado.
2. Muestra por pantalla de emparejamientos keypoint-mappoint, donde para proyectar los segundos sobre la imagen es necesario emplear los parámetros de distorsión. Al no disponer de los correctos, el punto del mapa aparece en una posición incorrecta, como se aprecia en la Figura 5a.
3. Ajuste manual de los parámetros de calibración hasta reducir la distancia entre el par keypoint-mappoint, logrando un ajuste como el de la Figura 5b.

Cuando se logra ajustar todos los puntos se asume que los parámetros de calibración son próximos a los reales. Se ha comprobado experimentalmente que sin llegar a ser ideal, esta solución proporciona un buen funcionamiento, permitiendo una correcta operación con todas las secuencias. En la Tabla 3 se presentan los parámetros de calibración obtenidos para los 4 endoscopios del estudio.



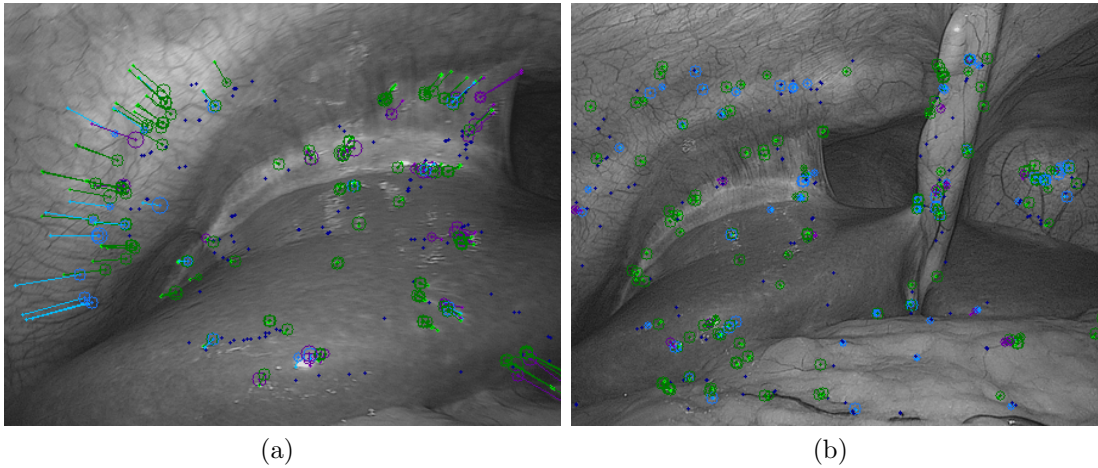


Figura 5: Para algunas secuencias la calibración se refina manualmente. Efecto en la imagen de la corrección: (a) sin corrección. (b) con corrección

Id	Tipo	Secuencias
1	Rígido	Cavidad abdominal: hígado, bazo, intestino(1) y bucles
2	Rígido	Cavidad abdominal: hígado y vesícula
3	Rígido	Cavidad abdominal: intestino(2)
4	Flexible	Gastroscopia: Esófago y estómago

Tabla 1: Endoscopios de los experimentos con el tipo de escena que muestra cada uno.

Origen	Resolución		Matriz de la cámara				Coeficientes de distorsión				
	Ancho	Alto	fx	fy	cx	cy	k1	k2	p1	p2	k3
IRCAD	1920	1080	925	918	981	576	-0.265	0.571	-0.006	0.002	-0.706
Propio	1920	1080	924	923	980	582	-0.249	0.451	-0.006	0.002	-0.453

Tabla 2: Verificación del software de calibración: los parámetros obtenidos son del mismo orden que los proporcionados por el IRCAD.

Endoscopio	Resolución		Matriz de la cámara				C. distorsión	
	Ancho	Alto	fx	fy	cx	cy	k1	k2
1	1920	1080	923	923	981	582	-0.226	0.256
	960	540	462	462	490	291	-0.226	0.256
	640	360	308	308	327	194	-0.226	0.256
2	960	540	450	450	514	328	-0.245	0.280
	640	360	300	300	343	219	-0.245	0.280
3	640	360	308	308	327	194	-0.200	0.280
4	768	576	403	414	386	277	-0.363	0.103
	480	360	336	345	241	173	-0.363	0.103

Tabla 3: Parámetros de calibración obtenidos para diferentes resoluciones, fijando en cero  $k_3$ ,  $p_1$  y  $p_2$ .



## 4.2. Sintonía

Para tratar escenas médicas con comportamiento no rígido, modificamos la forma de actuar del sistema al tratar los puntos, admitiendo ahora aquellos de leve movimiento, que des los cuales algunos serán identificados posteriormente como no rígidas solo las zonas que sufran deformación elevada, y sólo en ese caso no considerados en las estimaciones correspondientes. Para ello se introducen las siguientes modificaciones, cuyos valores alterados pueden encontrarse en la Tabla 4, junto con su situación sobre el código:

- En el proceso de tracking las modificaciones introducidas se basan en asumir que los keypoints detectados sobre la imagen van a sufrir cierto movimiento relativo, con lo que se amplía el área de búsqueda de emparejamientos en torno al punto del mapa predicho. Se considera un menor rango de escalas compatibles, porque hay regiones de la escena están muy próximas al plano de la imagen. Estas modificaciones pueden generar un mayor número de falsos positivos en los emparejamientos, para compensarlo se hace mas restrictiva la comparación entre descriptores ORB.
- En la gestión del mapa la primera modificación consiste en aumentar la distancia mínima a la que han de encontrarse dos keyframes para triangular nuevos puntos del mapa a partir de ellas, asegurando una mejor triangulación en una cavidad reducida. Con el fin de mapear de modo adecuado una envolvente rígida con la que funcione el sistema, se decide, contrariamente al caso del tracking, no aumentar la región de búsqueda de emparejamientos en torno a la línea epipolar, y ser más exigente en el error de reproyección permitido entre un punto del mapa y su keypoint asociado. Al mismo tiempo, se aumentan el paralaje mínimo exigido para llevar a cabo la triangulación, y se establece un emparejamiento más estricto entre descriptores.
- Tratando secuencias médicas se tiene la ventaja de que se dispone de conocimiento a priori sobre la geometría de la escena, y en particular el hecho de que en esta la no planitud será significativa. Por ello se fuerza la inicialización del sistema con el modelo de matriz fundamental, logrando así un arranque significativamente más rápido.
- Por último, en cuanto a la detección de ORBs: las imágenes a tratar tienen una gran cantidad de features sólo visibles en alta resolución, por ello se modifican los ficheros de Configuración (*Settings.yaml*), bajando el número de niveles de escala y el factor de cambio entre estos.

Descripción	Source	Línea	Antes	Ahora	Ud.
Tracking					
Área de búsqueda	ORBmatcher.cc	141	2.5/4.0	3.0/4.5	píxel
Invarianza de escala	Frame.cc	198	1/fe*	0.9/fe*	-
Comparación ORB	ORBmatcher.cc	44	50	45	bit
Gestión del mapa, creación de puntos					
Distancia entre keyframes	LocalMapping.cc	293	0.01	0.05	-
Distancia búsqueda g.epipolar	ORBmatcher.cc	166	3.84	3.84	píxel
Paralaje mínimo	LocalMapping.cc	354	1.1459	1.4035	Deg
Error de reproyección	LocalMapping.cc	413	5.991	0.5991	píxel
Comparación ORB	ORBmatcher.cc	44	50	45	bit
Inicialización					
Fundamental/Homografía	Initializer.cc	113	0.40	0.90	-
Settings					
Niveles de escala	Settings*.yaml	-	8	6	-
Factor entre niveles de escala	Settings*.yaml	-	1.2	1.1	-

Tabla 4: Sintonía realizada: descripción del parámetro modificado, archivo en el que se encuentra, línea en la que se ha cambiado el valor, valor original, valor modificado, unidades del parámetro. (\*fe:  $(\text{distancia al punto}) / (\text{factor de escala}) / (\text{factor de escala de keyframe asociado})$ )

A continuación se presenta una descripción de cada uno de los cambios mostrados en la Tabla 4:

- Área de búsqueda: región en torno al punto del mapa proyectado en la imagen, dentro de la que se buscan emparejamientos con keypoints de la imagen. Se aumenta el área puesto que los keypoints pueden haber sufrido un movimiento no rígido que los aleje de su posición predicha al asumirlos rígidos.
- Invarianza de escala: el parámetro indica la distancia entre el punto en 3D y el centro óptico de la cámara. Para un rango de distancias se asocia un punto con un determinado nivel de escala. Se amplía el extremo inferior del rango permitiendo puntos más cercanos a la cámara, abundantes en las imágenes tratadas.
- Comparación ORB (tracking): los keypoints se emparejan si la distancia de Hamming entre sus descriptores es inferior a un determinado valor. Se establece un emparejamiento más estricto para compensar el exceso de puntos susceptibles de emparejar debido al aumento del área de búsqueda. Comparamos con mas puntos pero solo emparejamos si la concordancia entre descriptores es alta.
- Distancia entre keyframes: se emplea el ratio entre la longitud de la línea que une los centros ópticos de dos keyframes, y la profundidad media de la escena.

Se aumenta el valor para buscar emparejamientos entre keyframes con mayor paralaje y que la posición 3D en la inicialización sea más precisa

- Distancia búsqueda epipolar: distancia de un keypoint a la línea epipolar, si es menor al umbral establecido se empareja el punto. Se mantiene el valor original.
- Paralaje mínimo: establece el ángulo que forman los rayos entre un punto en 3D y los centros ópticos de los keyframes que se han empleado para triangularlo. Se aumenta para asegurar una triangulación de calidad.
- Error de reproyección: Distancia entre la posición del punto del mapa reproyectado en los keyframes que lo han triangulado, y los keypoints que se han empleado para ello. Se reduce para asegurar que solo pasan al mapa puntos triangulados con precisión –idealmente debería ser cero.
- Comparación ORB (gestión del mapa): razonamiento similar al tracking, se favorece que los emparejamientos sean correctos.
- Fundamental/Homografía: El valor representa el ratio entre el error obtenido al emplear matriz fundamental y homografía. Se aumenta para forzar la inicialización con matriz fundamental.
- Niveles de escala: Número de niveles de escala en los que se buscan ORBs de la imagen. Se reduce, puesto que se tiene gran cantidad de puntos de interés de alta resolución.
- Factor de escala: Factor de cambio entre los niveles de escala: Se reduce minimizando aun más el cambio de resolución.

## 5. Experimentos

Las operaciones de endoscopia se clasifican según el orificio por el que se introduce el endoscopio y la cavidad observada: en nuestro estudio se tratan escenas propias de laparoscopia, que exploran la cavidad abdominal, y gastroscopia, que hacen lo propio en el aparato digestivo superior.

Dado que la escena vista es radicalmente distinta en una u otra operación, el estudio se enfoca en analizar los procesos de emparejamiento y de gestión del mapa en cada una de ellas; prestando especial atención a la forma en que se identifican como no rígidas porciones de la escena. Para ello nos centramos en el estudio de los puntos del mapa y de los emparejamientos con estos, no prestando atención al tratamiento de keyframes, pues depende mayormente del movimiento de la cámara, no del de los puntos.

Se ha modificado el código para para mostrar por pantalla las etapas del proceso de tracking y de gestión del mapa, estudiando todas las etapas que realiza el programa al tratar los puntos.

Los siguientes párrafos hacen las veces de leyenda de lo que se mostrará en las imágenes, pudiendo encontrar información más detallada en el Anexo A.

Siguiendo el funcionamiento del sistema expuesto en la sección 3, la evaluación del tracking clasifica los puntos del mapa según sean:

- Puntos predichos y emparejados con un ORB de la imagen ( $\odot$ ).
- Puntos predichos pero no emparejados por falta de Keypoints próximos en la imagen ( $\bullet$ ).
- Emparejamientos con puntos del mapa que aún están en etapa infantil ( $\odot$ ).
- Puntos emparejados que son descartados por la función de ajuste robusto al no cumplir la condición de rigidez ( $\odot$ ).

En cuanto a gestión del mapa se estudia el proceso de creación de puntos:

- Cuantificamos los emparejamientos entre keyframes y cuantos de estos dan puntos del mapa, presentando esta información en tablas con estadísticas.
- Estudio de los puntos que cumplen con la geometría epipolar pero son descartados por paralaje insuficiente ( $\square$ ), error de reproyección ( $\square$ ), inconsistencia en la escala ( $\square$ ), o profundidad negativa ( $\square$ ).
- Estudio de la etapa mortalidad infantil de los nuevos puntos, evaluando el borrado que se produce durante estricta evaluación en los 3 keyframes próximos a su creación, identificando la edad del punto con diferente color según si tiene 1 keyframe ( $\odot$ ), 2 keyframe ( $\odot$ ), 3 keyframes ( $\odot$ ), o si se ha vuelto punto permanente ( $\bullet$ ).
- Se busca conocer la razón de que un punto no llegue a ser permanente.

Se analizan 19 secuencias tomadas in vivo durante diferentes operaciones, en diferentes animales y con diferentes endoscopios, garantizando la fiabilidad de los resultados obtenidos. Cada escena vista tiene diferente comportamiento y el estudio se organiza según el tipo de endoscopia realizada (laparoscopia o gastroscopia) y según el área observada.

Las secuencias se reproducen en resolución 640x360 píxeles, realizando varias ejecuciones de cada una. Para cada escena se lleva a cabo un estudio visual de la distribución

de puntos sobre la imagen, y se extraen valores numéricos cuantificando el estado en que se encuentran los puntos del mapa. Se presenta un resumen de los valores obtenidos, pudiendo encontrar datos adicionales para todas las secuencias en el Anexo III.

## 5.1. Cavidad abdominal: hígado y bazo

Se tienen 4 secuencias en las que se muestran hígado, bazo, y un background compuesto de pared abdominal y un fragmento de intestino no relevante en el estudio de estas secuencias. Se divide la escena en 3 regiones según el órgano observado, pues cada uno presenta diferente comportamiento: el hígado es mayormente rígido pero con movimiento localizado en su contorno, el bazo puede considerarse un conjunto rígido, y la pared abdominal sufre movimiento localizado causado por los latidos del corazón.

Estas son las escenas de mayor calidad disponibles, con una envolvente rígida definida y zonas no rígidas localizadas. Además las secuencias presentan total ausencia de ruido, pocos reflejos, y se dispone de una calibración precisa corroborada por los datos cedidos por el IRCAD.

Se proporciona un vídeo que muestra el funcionamiento del sistema al procesar una de estas secuencias: <https://youtu.be/TIpkP5N3DOI>

### 5.1.1. Tracking

Se reproducen las 4 secuencias de forma consecutiva logrando un funcionamiento estable y un mapa de calidad, y efectuando relocalizaciones correctas cuando es necesario. En la Tabla 5 se presentan estadísticas del estado en que se encuentran los puntos del mapa que intervienen en el tracking.

Estado del punto del mapa	Media	Mediana	% Predicho	Rango
Predicho en la imagen	557	573	-	296 - 815
Emparejado con un ORB	165	154	28	70 - 322
No emparejado: sin un ORB cercano	326	335	61	151 - 490
No emparejado: sin un ORB similar	1	0	0	0 - 1
No cumple la condición de rigidez	66	64	12	22 - 108

Tabla 5: Tracking en hígado y bazo. El porcentaje está referido al total de puntos predichos en la imagen (valor de mediana).

Resalta que en cada ejecución se tienen una gran cantidad de puntos del mapa predichos en el frame que no pueden ser emparejados porque no tienen KeyPoints cercanos con los que intentarlo, notéense los puntos azul oscuro en la Figura 6a.

En cuanto al estudio del comportamiento no rígido, toda la escena tiene un leve movimiento, sin embargo el sistema es capaz de extraer una envolvente rígida, con la

que realiza el tracking y genera un mapa definido, Figura 6b. En las áreas de la escena que presentan un comportamiento no rígido mas intenso, como el contorno del hígado, el emparejamiento no se lleva a cabo cuando la región no se encuentra en la posición original para la que se crearon los puntos del mapa que la definen, como puede verse en las Figuras 6c y 6d.

### 5.1.2. Gestión del mapa

En las Tablas 6 y 7 se presentan estadísticas en cuanto a la creación y borrado de puntos del mapa.

	Media	Mediana	% Epipolar	Rango
ORBs disponibles para emparejar	1200	1209	-	1053-1301
Cumplen geometría epipolar	216	221	-	55-361
Descartado por profundidad negativa	29	25	13	5-71
Descartado por paralaje insuficiente	17	11	6	0-59
Descartado por error de reproyección	34	32	16	7-67
Descartado por escala inconsistente	2	0	0	0-10
Descartado por no cumplir rigidez	67	63	32	6-141
Puntos creados	67	63	32	15-141

Tabla 6: Gestión del mapa en hígado y bazo: creación de puntos. El porcentaje está referido al total de puntos que cumplen con la geometría epipolar.

Estado del punto del mapa	Media	Mediana	% Total	Rango
Eliminados por no ser emparejables	42	37	37	6-92
Eliminados por no encontrarse en keyframes	6	4	4	0-16
Eliminados por no cumplir rigidez	61	60	59	17-108

Tabla 7: Gestión del mapa en hígado y bazo: borrado de puntos. El porcentaje está referido respecto del total de puntos descartados: los no emparejables, mas los que no se encuentran en keyframes, mas los que no cumplen rigidez

En la Figura 6e se presentan las comprobaciones realizadas sobre los puntos que cumplen la geometría epipolar. Una vez superadas se entra en la etapa de mortalidad infantil de los puntos del mapa durante los 3 keyframes siguientes a su nacimiento, donde se tiene que el mayor descarte de puntos es causado porque el tracking no es capaz de emparejar el punto del mapa en al menos el 25% de los frames en los que este punto es predicho. Se considera que esto es causado por la ausencia de un ORB próximo con el que emparejarse. En la Figura 6f puede apreciarse que gran parte de los puntos del mapa no sobreviven.

Una vez el punto entra en el mapa de forma permanente este solo puede ser borrado si se elimina su keyframe asociado o si el ajuste robusto detecta que no se corresponde con un punto rígido; se ha encontrado que este borrado es mínimo si el sistema se encuentran en funcionamiento estable, habiendo pasado un tiempo desde la inicialización.

El problema encontrado en el tracking, por el que no existen puntos lo bastante próximos para ser susceptibles de emparejarse no afecta en esta ocasión, siendo la densidad de puntos muy superior al buscar emparejamientos entre varios keyframes.

## 5.2. Cavity abdominal: hígado y vesícula biliar

Se dispone de 4 secuencias de exploración de la zona abdominal previas a una operación de colecistectomía, esto es, la extirpación de la vesícula biliar, en ellas se recorre el entorno pasando varias veces sobre el mismo área con movimientos de traslación, ideales para el funcionamiento del sistema.

Al igual que en el apartado anterior, la escena presenta una combinación de áreas con componente rígido y no rígido, a lo largo de una cavidad amplia y en unas imágenes de gran calidad, con ausencia de ruido y pocos reflejos, mostradas en la Figura 7a.

En este caso se encuentra que el mapa generado es incluso de mayor calidad que el anterior, entendiéndose por calidad un mapa definido en el que se puede identificar la escena, Figura 7b, debido sin duda a que predominan los movimientos de traslación sobre los de rotación. La evaluación devuelve conclusiones similares a las anteriores, teniendo ausencia de ORBs cercanos a los puntos predichos del mapa, y una identificación de la parte no rígida de la escena llevada a cabo por la función de ajuste robusto, que se aprecia en las Figuras 7c y 7d.

En las Figuras 7e y 7f se observa un comportamiento similar a la escena anterior en cuanto a la creación de puntos: con diversos rechazos tras confirmar geometría epipolar, y con pocos puntos del mapa que superan la evaluación inicial, nuevamente por no ser emparejados en el tracking.

Pueden consultarse valores estadísticos análogos a los del apartado anterior en la sección 2 del Anexo C. Además se proporciona un vídeo que muestra el funcionamiento del sistema al procesar una de estas secuencias: <https://youtu.be/4PTN-CzCqUo>

## 5.3. Cavity abdominal: intestino

Se dispone de dos juegos de secuencias que muestran la zona intestinal, tomados con diferentes endoscopios (1 y 3) en diferentes animales. En ambas se tiene una imagen de gran calidad, con ausencia de ruido, de reflejos importantes, y con una calibración precisa.

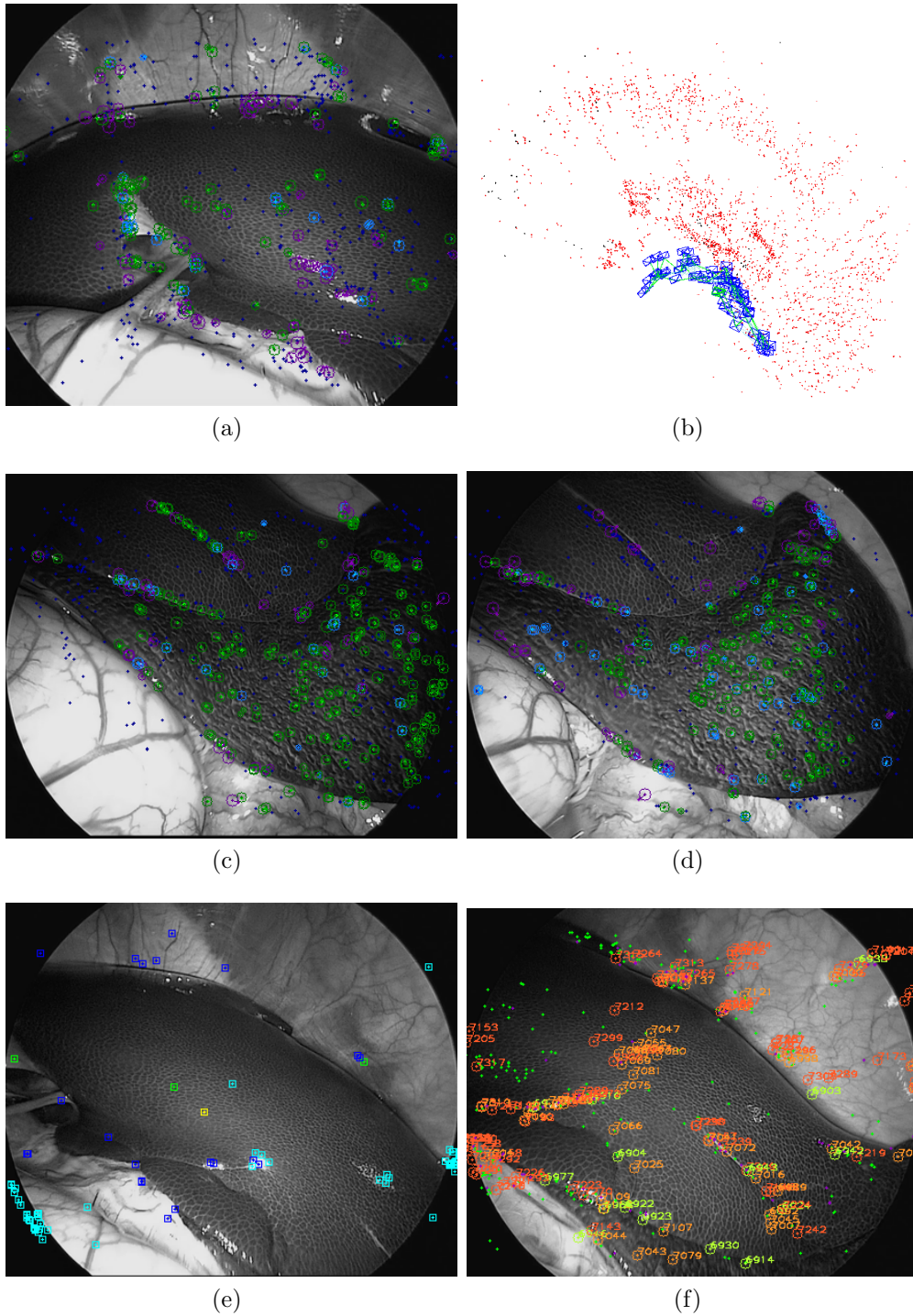


Figura 6: Tracking en la laparoscopia del hígado y bazo: (a) Gran cantidad de puntos del mapa ( $\bullet$ ) no son emparejados por no existir ORBs cercanos en la imagen. (b) Vista general del mapa 3D y de la trayectoria de la punta del endoscopio. Algunos puntos de la escena son no rígidos, cuando son consistentes con la envolvente rígida se marcan como inliers ( $\odot$ )(c), cuando no son consistentes se marcan como outliers ( $\ominus$ )(d). Inicialización de nuevos puntos en el mapa: (e) Algunos puntos del mapa que cumplen geometría epipolar se descartan por paralaje insuficiente ( $\blacksquare$ ), error de reproyección ( $\square$ ), inconsistencia en la escala ( $\blacksquare$ ), o profundidad negativa ( $\square$ ). (f) Inicialmente se crean muchos puntos del mapa ( $\odot$ ) pero pocos sobreviven hasta la tercera keyframe ( $\ominus$ ).



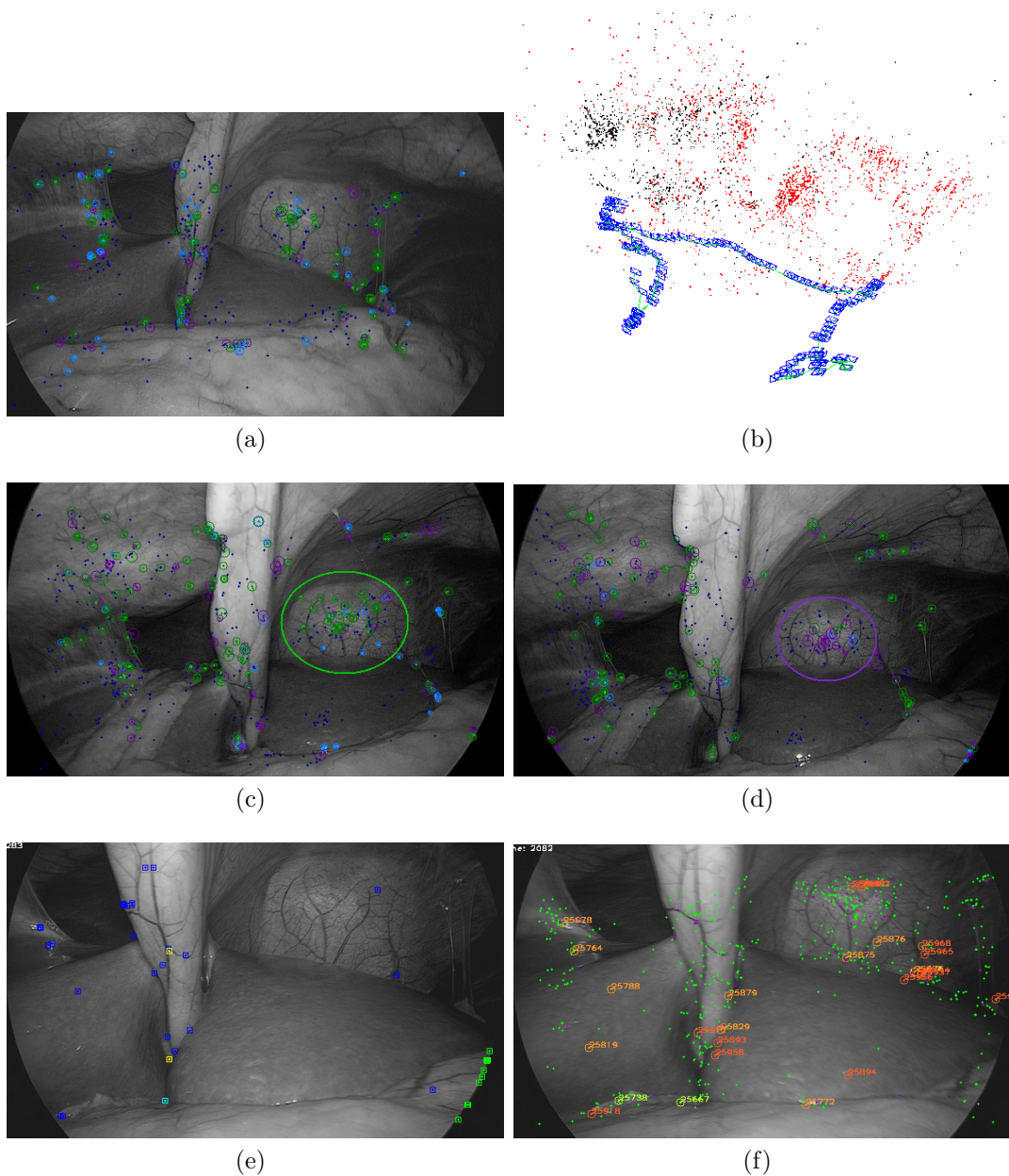


Figura 7: Laparoscopia en hígado y vesícula biliar. (a) Tracking en la imagen. (b) Mapa generado, notese el predominio de movimientos de traslación. (c) Los puntos señalados debido a la respiración son no rígidos, cuando son consistentes con la envolvente rígida se marcan como inliers ( $\odot$ ). (d) La respiración deforma la pared abdominal, sus puntos ya no cumplen la condición de rigidez y son marcados como outliers ( $\odot$ ). Inicialización de nuevos puntos en el mapa: (e) Algunos puntos del mapa que cumplen geometría epipolar se descartan por paralaje insuficiente ( $\square$ ), error de reproyección ( $\square$ ), inconsistencia en la escala ( $\square$ ), o profundidad negativa ( $\square$ ). (f) Inicialmente se crean muchos puntos del mapa ( $\odot$ ) pero pocos sobreviven hasta la tercera keyframe ( $\odot$ ).

Las secuencias tomadas con el endoscopio 1 presentan mayor dificultad pues muestran una escena de elevado movimiento, a pesar de ello el sistema inicializa sin problemas y realiza una adecuada reconstrucción de la escena durante un tiempo.

Durante el proceso de tracking, ante la presencia de áreas con movimiento el sistema responde anclando el tracking en una porción delimitada de intestino, haciendo de esta su referencia rígida, permitiendo el funcionamiento en estas escenas de elevada dificultad.

Se logran emparejar puntos del resto de la imagen si estos no sufren movimiento relativo respecto de la referencia; siendo la función de ajuste robusto la que descarta zonas que no cumplen con la condición de rigidez, presentamos un ejemplo de este funcionamiento en la Figura 8a.

A pesar de la adecuada reconstrucción inicial el funcionamiento correcto no se extiende en el tiempo, conforme avanza la secuencia se observa una deformación del mapa que acaba causando un fallo del tracking y dificultad para efectuar relocation.

En contraposición las secuencias del endoscopio 3 muestran un intestino con elevada rigidez, lo que las hace muy fáciles de tratar a pesar de para ellas solo se dispone de calibración manual.

La combinación de elevada rigidez con un movimiento de cámara de traslación durante toda la secuencia permite la inicialización más rápida encontrada en todas las secuencias, además de facilitar el tracking y la generación del mapa, encontrando un buen funcionamiento que se muestra en las Figuras 8b y 8c.

En ambos grupos de secuencias se repite el problema reseñado anteriormente, una falta de emparejamientos causada por la ausencia de ORBs próximos a los puntos predichos del mapa.

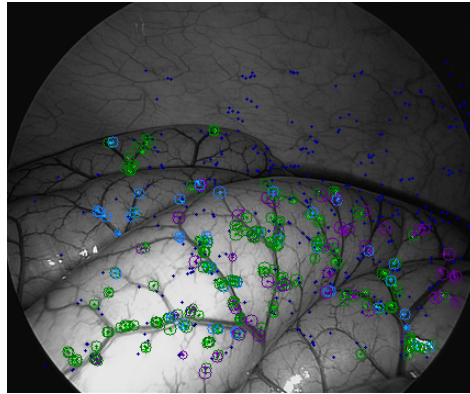
Se proporciona un vídeo que muestra el funcionamiento del sistema al procesar estas secuencias, mostramos unos segundos de cada uno de los endoscopios para permitir apreciar el efecto de la diferencia de rigidez en la reconstrucción: <https://youtu.be/eSm8Wv4qIo4>

#### **5.4. Endoscopia digestiva alta: esófago y estómago**

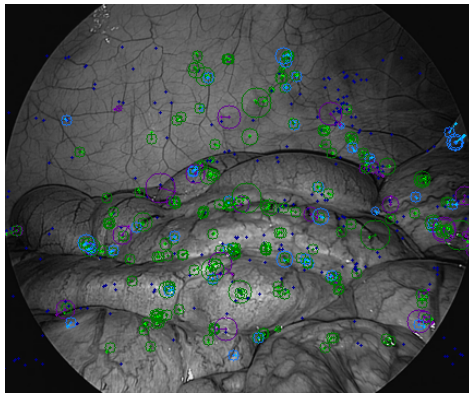
Se dispone de secuencias de gastroscopia radicalmente distintas a las anteriores, en ellas el instrumento es introducido por la boca del animal, recorriendo rápidamente el esófago para acabar llevando a cabo una exploración exhaustiva del interior del estómago.

Las secuencias están grabadas con un endoscopio flexible y son de baja calidad, con gran cantidad de ruido, lo que ha hecho difícil lograr una calibración precisa.

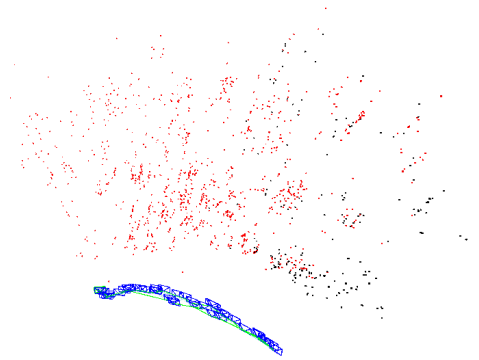
La secuencia se divide en dos etapas:



(a)



(b)



(c)

Figura 8: Laparoscopia en zona intestinal: (a) Endoscopio 1: el sistema identifica como rígida una porción de la escena, el tubo mayor del intestino, y realiza el tracking desde este, ignorando la mayor parte del resto de la imagen. (b) Endoscopio 3: en escena esta el comportamiento no rígido es significativamente menor, el tracking empareja puntos en toda la imagen. (c) Endoscopio 3: mapa generado: distribución de puntos especialmente uniforme por la combinación de movimiento de traslación y alta rigidez.

- La primera muestra el avance del endoscopio por el esófago, el movimiento por una cavidad angosta y con gran cantidad de oclusiones y reflejos dificulta enormemente el funcionamiento del sistema, Figura 9a.

Eventualmente se logra inicializar y realizar una reconstrucción de la escena, Figura 9b, pero el tracking dura pocos segundos, siendo entonces imposible relocalizar la cámara.

- Al atravesar el cardias y entrar en el estómago el cambio a un área más amplia permite inicializar correctamente cuando el endoscopio realiza un movimiento de traslación respecto a la pared estomacal.

La escena vista, Figura 9c, presenta una gran cantidad de reflejos, y una superficie con pocas features de calidad (alto gradiente en dos direcciones), lo que impide que el emparejamiento de un mismo punto se prolongue en el tiempo.

Esta ausencia de emparejamientos eficaces causa que muy pocos puntos del mapa superen la evaluación inicial, con lo que no se logra una reconstrucción eficaz desde la que llevar a cabo el tracking, Figura 9d.

Se presenta un vídeo que muestra el funcionamiento del sistema con estas secuencias: <https://youtu.be/IIxYtJhmTkA>

## 5.5. Bucles en la cavidad abdominal

Para completar los experimentos se prueban dos últimas secuencias en las que la cámara sigue una trayectoria circular, realizando una exploración completa de la cavidad abdominal. Se busca comprobar si el sistema es capaz de mantener el tracking durante todo el recorrido llegando a efectuar un cierre de bucle.

Se tiene un comienzo prometedor con una precisa reconstrucción inicial, Figura 10a, sin embargo conforme la cámara avanza, el componente no rígido de la escena causa que se acumule deformación no rígida en el mapa generado, Figura 10b. Una vez se completa el recorrido, la falta de rigidez de la escena causa que el segmento de mapa que esta frente a la cámara sea diferente al generado para el mismo área al comienzo del bucle, lo que impide la detección del cierre de bucle, que requiere tanto similaridad entre los keyframes (cumplido), como que el mapa que se ha formado en ambos extremos no sufra deformación no rígida entre uno y otro (no cumple).

La imprecisión en la reconstrucción surge al intentar mapear órganos independientes en una misma secuencia, siendo que existe movimiento relativo entre ellos, el error acumulado es mas evidente conforme nos alejamos del punto de inicio.

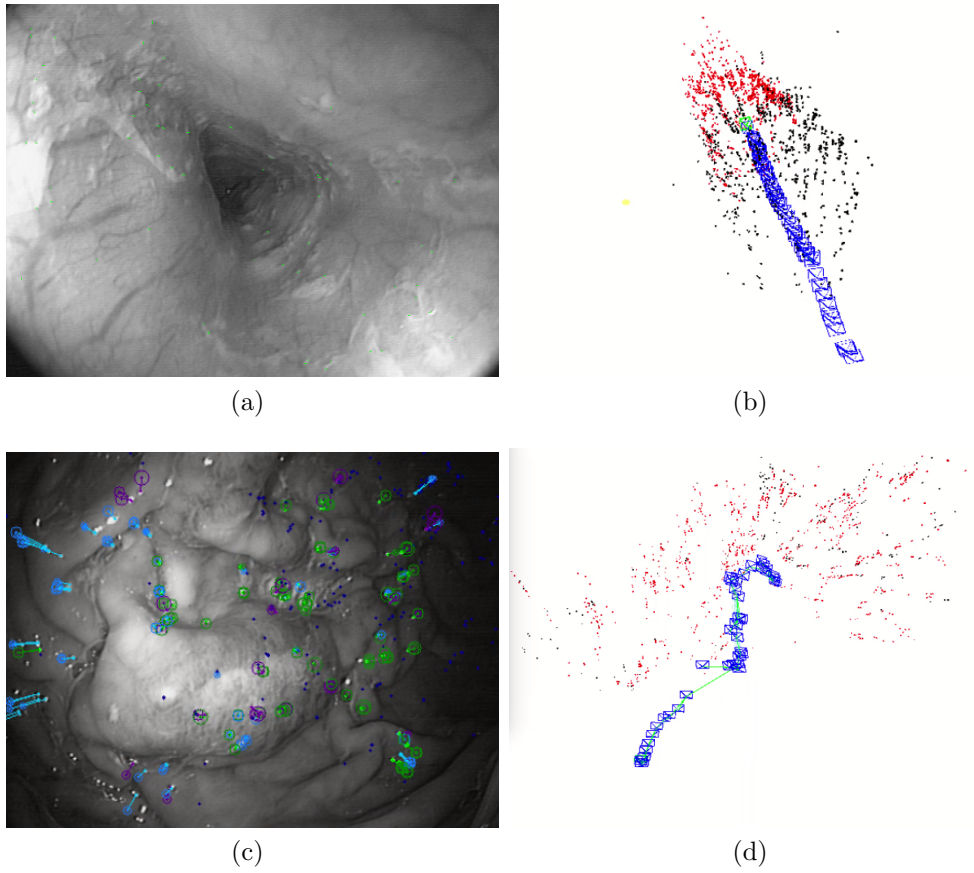


Figura 9: Gastroscopia: (a) Vista típica del esófago: zona húmeda en la que se esperan importantes oclusiones y reflejos. (b) Mapa del esófago: buena definición del movimiento de avance por el interior de un cuerpo cilíndrico. (c) Escena del estómago: muchos ORBs situados sobre los reflejos. (d) Mapa generado en el estómago, la trayectoria seguida por la punta del endoscopio ha acumulado una deriva importante.

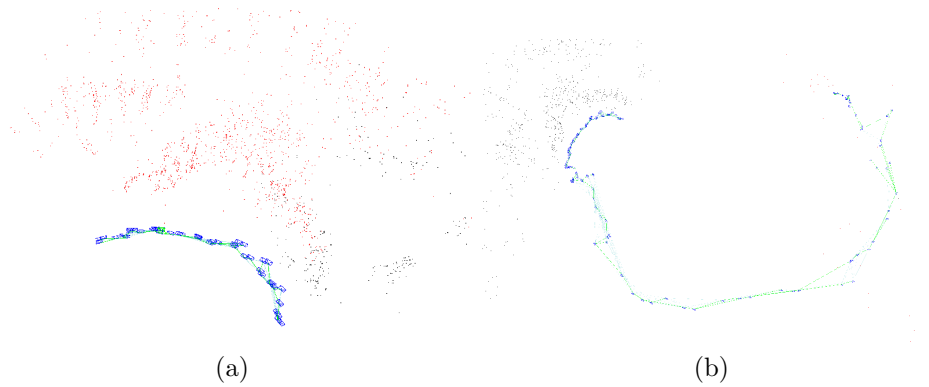


Figura 10: Bucles en la cavidad abdominal: (a) el mapa inicial es compacto y preciso. (b) Al continuar el recorrido se acumula un error importante, los dos extremos superiores corresponden al mismo punto de la escena, la deformación del entorno impide cerrar el bucle.

## 6. Resultados

Tras la evaluación de 19 secuencias tomadas in-vivo durante operaciones de endoscopia médica, se ha confirmado que el sistema ORBSLAM tiene capacidad de funcionar con secuencias de este tipo, encontrando que el tipo de escena tratada afecta en gran medida al funcionamiento del sistema:

- Durante operaciones de laparoscopia que exploran la cavidad abdominal, se tiene un funcionamiento correcto, pudiendo construir un mapa de la escena y utilizar este para localizar la cámara sin error.

Las zonas de alto componente no rígido de la escena son identificadas por la función de ajuste robusto, y descartadas para el tracking y la creación de puntos. Si el área no rígida se encuentra localizada, su presencia no presenta un impedimento al funcionamiento del sistema. Solo en escenas en la que todas sus partes sufren movimiento relativo entre sí (intestino) se encuentra una dificultad al mantener el tracking.

- En operaciones de gastroscopia se tiene poco espacio de maniobra, durante el avance por el esófago se presentan frecuentes oclusiones y movimientos bruscos del endoscopio; y una vez en el estómago la gran cantidad de reflejos y la ausencia de features de calidad en la escena impiden un funcionamiento correcto del sistema.

La función de ajuste robusto reconoce las áreas de fuerte componente no rígido y las diferencia del resto de la escena; el resto de puntos de poco movimiento relativo son considerados como rígidos y tratados adecuadamente.

Se ha encontrado un problema común a todas las escenas en cuanto al emparejamiento de puntos del mapa, siendo que para gran parte de los puntos no hay ORBs cercanos con los que intentar emparejarlos.

### 6.1. Discusión

Tras las numerosas pruebas, se ha encontrado que el sistema ORBSLAM, diseñado para robótica, presenta un rendimiento elevado al tratar escenas con componente no rígido para las que en principio no ha sido diseñado. Si bien el funcionamiento correcto esta condicionado hay un amplio espacio para mejorar su rendimiento.

Para las escenas de laparoscopia, se ha continuado el trabajo de [10, 11], evaluando un nuevo sistema que es capaz de hacer mapas de mucho más tamaño, mayor densidad de puntos y mayor robustez en la operación.

Los problemas encontrados al tratar escenas de gastroscopia se deben a la naturaleza de las secuencias: poco espacio, pocas features de calidad, reflejos, oclusiones,

movimiento brusco del endoscopio. Los resultados obtenidos en esta sección concuerdan con los presentados en [13]. Siendo que estas secuencias presentan interés en operaciones de detección de cáncer de esófago [14], se propone estudiar las ideas propuestas en [15]. Reseñar que durante estas operaciones se emplean endoscopios flexibles, los cuales presentan mayor distorsión que los empleados en laparoscopia, lo que añade una dificultad adicional; en cualquier caso, disponemos de un reducido número de secuencias de gastroscopia, con una calibración pobre, por lo que se recomienda realizar más experimentos de forma controlada para evaluar de forma fiable las posibilidades de tratamiento en estas secuencias.

En vistas a continuar la investigación, creemos necesario implementar un enfoque multidisciplinar de las operaciones, en las que el personal médico colabore con el personal de investigación favoreciendo una toma adecuada de imágenes, haciendo un esfuerzo por mantener fijos el zoom de la cámara o el enfoque, que ayudara en el funcionamiento correcto del sistema, también sería interesante disponer de soluciones de referencia tomadas con sensores tomográficos tales como el CT.

Cabe reseñar que todas las escenas evaluadas se corresponden con endoscopia exploratoria, si se avanza en la operación y se comienza la cirugía endoscópica, la escena hasta entonces reconstruida cambiará radicalmente, en principio impidiendo el funcionamiento con la información de la escena recopilada anteriormente.

## 6.2. Conclusiones

En el trabajo se ha evaluado la viabilidad de emplear un sistema de Visual SLAM, el ORBSLAM, en escenas de endoscopia médica. Siendo un sistema diseñado para escenas rígidas propias de robótica, se ha sintonizado el código para lograr el tratamiento de escenas no rígidas propias de secuencias de endoscopia médica.

Se ha llevado a cabo una evaluación del sistema con 19 secuencias tomadas in-vivo durante operaciones de endoscopia médica, tras lo que se concluye que el sistema ORBSLAM tiene capacidad de operar con este tipo de escenas, con rendimiento alto durante procedimientos de laparoscopia abdominal; y se recomienda profundizar en su aplicación con las modificaciones que se presentan la siguiente sección.

Se encuentra fundamental el input que se proporciona al sistema, pues la presencia de ruido o reflejos, gastroscopia, presenta más problemas al funcionamiento que la existencia de áreas con componente no rígido, laparoscopia (intestino).

La existencia de una calibración adecuada juega un papel determinante. El diseño de nuevos experimentos debe incluir la adquisición de imágenes de calibración de calidad: posiciones que proporcionen buen condicionamiento, imágenes llenas de puntos del patrón y posición y tamaño del patrón similares a los del área de trabajo.

### 6.3. Trabajo futuro

En base al análisis de los resultados obtenidos se proponen varias líneas de trabajo a seguir, actuando en las diferentes etapas del proceso de funcionamiento:

**Extracción de features** .- Se ha reseñado que en el proceso de tracking se produce una falta de emparejamientos debido a que no se han extraído ORBs cerca de los puntos del mapa predichos en el frame. Se propone profundizar en la implementación de detector de puntos de interés para conseguir que dispare en las regiones de búsqueda de los puntos del mapa reproyectado.

También podría considerarse la eliminación de reflejos.

**Segmentación** .- Durante la cirugía el área de interés se limita a una región concreta; se considera importante implementar una segmentación de los órganos de interés respecto del resto de la escena. Continuamente el sistema interpreta como rígida una región de la escena a partir de la cual lleva a cabo el tracking, la existencia de solo un órgano en la imagen forzaría a construir una envolvente rígida de este órgano, facilitando la identificación y tratamiento de su parte no rígida.

**Tratamiento de puntos no rígidos** .- La función de ajuste robusto reconoce las áreas no rígidas de la escena, actualmente termina eliminándolos. Se propone desactivar el borrado de los puntos del mapa pertenecientes a estas áreas, con el objetivo de poder almacenar el tracking de los puntos no rígidos para poder hacer un post-proceso con modelos de SLAM no rígido de la escena.

**Calibración** .- Aplicar métodos de autocalibración.



## 7. Referencias

- [1] Davison A.J., Reid I., Molton N.D., Stasse O., *MonoSLAM: Real-time single camera SLAM*, IEEE Transactions on Robotics, vol. 29, no. 6, pp. 1052-1067, 2007.
- [2] Cheeseman P., Smith R., Self M., *A stochastic map for uncertain spatial relationships*, 4th International Symposium on Robotic Research, 1987.
- [3] Civera J., Davison A.J., Montiel J.M.M., *Interacting multiple model monocular SLAM*, IEEE International Conference on Robotics and Automation, pp. 3704-3709, 2008.
- [4] Civera J., Grasa O.G., Davison A.J., Montiel J.M.M., *1-Point RANSAC for Extended Kalman Filtering: Application to Real-Time Structure from Motion and Visual Odometry*, IEEE Journal of Field Robotics, vol. 27, no. 5, pp. 609-631, 2010.
- [5] Klein G., Murray D., *Parallel Tracking and Mapping for Small AR Workspaces*, IEEE and ACM International Symposium on Mixed and Augmented Reality, 2007.
- [6] Mikhail E. M., Bethel J. S., McGlone J. C., *Introduction to modern photogrammetry*, vol. 1, John Wiley & Sons Inc, 2001.
- [7] Raul Mur-Artal, J. M. M. Montiel, Juan D. Tardós, *ORB-SLAM: A Versatile and Accurate Monocular SLAM System*, IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, 2015.
- [8] Rublee E., Rabaud K. K., Bradski G., *ORB: an efficient alternative to SIFT or SURF*, IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, pp. 2564-2571, 2011.
- [9] Lowe D. G., *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.
- [10] García O.G., Bernal E., Casado S., Gil I., Montiel J. M. M., *Visual SLAM for Hand-Held Monocular Endoscope*, IEEE Transactions on Medical Imaging, vol. 33(1), pp. 135-146, 2014.
- [11] García O.G., Civera J., Güemes A., Muñoz V., Montiel J.M.M., *Real-time 3D Modeling from Endoscope Image Sequences*, IEEE International Conference on Robotics and Automation (ICRA2009), Kobe (Japan), 2009.

- [12] Hartley R., Zisserman A., *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [13] Castaneda V., Atasoy S., Mateus D., Navab N.O., Meining A., *Reconstructing the Esophagus Surface from Endoscopic Image Sequences*, Bavarian Conference on Bio-Medical Engineering, Munich, Germany, 2009.
- [14] Sharma P., Bansal A., Mathur S., Wani S., Cherian R., McGregor D., Higbee A., Hall S., Weston A. *The utility of a novel narrow band imaging endoscopy system in patients with Barrett's esophagus*, *Gastrointestinal Endoscopy*, 64, 167-175, 2006.
- [15] Zhou J., Das A., Li F., Li B., *Circular generalized cylinder fitting for 3d reconstruction in endoscopic imaging based on mrf*. 1-8, 2008.

# Anexos

## A. Leyenda de las ventanas

Se ha modificado el código añadiendo ventanas adicionales para mostrar por pantalla el estado de los puntos del mapa y los emparejamientos logrados, o la ausencia de estos. El objeto del Anexo A del Trabajo de Fin de Grado es el de detallar que se muestra en cada ventana, con una leyenda para su interpretación.

### A.1. Ventana 1: Tracking

Se muestran los puntos del mapa predichos sobre el frame, con diferente color según su estado. Si se logra emparejar el punto del mapa con un keypoint del frame, se dibuja el par emparejado (mappoint-keypoint) unido con una línea del mismo color. Para cada punto del mapa se dibuja un círculo que señala la región en la que se buscan emparejamientos.

En las Tablas 8, 9 y 10, se describe que representa cada color dibujado.

### A.2. Ventana 2: Gestión del mapa (edad de los puntos)

Se muestran los puntos del mapa indicando su edad, esto es, el número de keyframes que han pasado desde su creación, diferenciándolos con diferente color.

En la Tabla 11 se describe que representa cada color dibujado.

### A.3. Ventana 3: Gestión del mapa (borrado de puntos)

Se muestra el proceso de creación de puntos del mapa, dibujando los emparejamientos que se producen entre keyframes, y la razón por la que se descartan estos emparejamientos.

En la Tabla 12 se describe que representa cada color dibujado.

Verde oscuro	Punto del mapa reproyectado sobre el frame actual
Verde claro	ORB con el que se ha emparejado el punto del mapa

Tabla 8: Puntos del mapa emparejados en el tracking

Azul oscuro	Punto del mapa reproyectado sin keypoint cercano
Azul medio	Punto del mapa emparejado que acaba descartándose
Azul claro	ORB con el que se ha emparejado el punto del mapa

Tabla 9: Puntos del mapa no emparejados

Morado oscuro	Punto del mapa reproyectado
Morado claro	ORB con el que se ha emparejado el punto del mapa

Tabla 10: Puntos del mapa descartados por la función de ajuste robusto

Naranja	0 keyframes, punto recién creado
Amarillo	1 keyframe
Limón	2 keyframes
Verde	3 keyframes, acabada la evaluación inicial
Morado	Marcado como outlier por la función de ajuste robusto

Tabla 11: Edad de los puntos del mapa

Cyan	Profundidad negativa del punto triangulado en alguna keyframe
Verde	Paralaje insuficiente
Azul	Mucho error de reproyección
Amarillo	Inconsistencia en la escala

Tabla 12: Razón del fallo en la creación de puntos del mapa

## B. Secuencias del estudio

El objeto Anexo B del Trabajo de Fin de Grado es el de listar y describir las secuencias de vídeo de las que se dispone. Como se ha señalado en la sección 4, estas se organizan por endoscopio y por escena vista.

### B.1. Endoscopio 1

Nombre original: 6080HD\*

Resolución de origen: 1920x1080

Resoluciones de trabajo: 640x360(lr) y 960x540(mr)

Tipo de endoscopia: laparoscopia.

Descripción: Se muestra la cavidad abdominal, con secuencias centradas independientemente en el hígado, el bazo, el intestino, y llevando a cabo dos bucles de la cavidad. La escena tiene componentes no rígidos localizados en los contornos del hígado y en tramos de intestino.

En la Tabla 13 se listan las secuencias correspondientes al endoscopio 1. El nombre de columna *Bag* hace referencia al nombre de la bolsa de imágenes creada.

Bag	Área mostrada	Movimiento	Duración
*6080HDseq1.bag	Hígado, bazo	Bamboleo	110.9 s
*6080HDseq2.bag	Hígado, bazo	Bamboleo	33.1 s
*6080HDseq3.bag	-	Bucle	35.8 s
*6080HDseq4.bag	Hígado	Estática	36.0 s
*6080HDseq5.bag	-	Bucle	104.0 s
*6080HDseq6.bag	Intestino	Bamboleo	7.7 s
*6080HDseq7.bag	Intestino, bazo	Bamboleo	21.7 s
*6080HDseq8.bag	Hígado	Estática	110.9 s

Tabla 13: Endoscopio 1: cavidad abdominal

### B.2. Endoscopio 2

Nombre original: Luc\*

Resolución de origen: 1920x1080

Resolución de trabajo: 640x360(lr) y 960x540(mr)

Tipo de endoscopia: laparoscopia, colecistectomía. Descripción: Vesícula biliar y su entorno. Escenas de gran calidad con áreas no rígidas localizadas.

En la Tabla 14 se listan las secuencias correspondientes al endoscopio 3.

Bag	Área mostrada	Movimiento	Duración
*LUCseq1.bag	Entorno de la vesícula	Bamboleo	84.5 s
*LUCseq2.bag	Entorno de la vesícula	Bamboleo	73.1 s
*LUCseq3.bag	Entorno de la vesícula	Bamboleo	184.6 s
*LUCseq4.bag	Entorno de la vesícula	Bamboleo	40.1 s

Tabla 14: Endoscopio 3: cavidad abdominal, colecistectomía

### B.3. Endoscopio 3

Nombre original: seq\_Stephane\_Nicolo.pano\*

Resolución de origen: 640x360

Resolución de trabajo: 640x360(lr)

Tipo de endoscopia: laparoscopia.

Descripción: Escenas del intestino. Escena de gran rigidez y en vídeos de alta calidad.

En la Tabla 15 se listan las secuencias correspondientes al endoscopio 2.

Bag	Área mostrada	Movimiento	Duración
lrSNPseq1.bag	Intestino	Bamboleo	76.9 s
lrSNPseq2.bag	Intestino	Bamboleo	10.4 s
lrSNPseq3.bag	Intestino	Bamboleo	20.9 s

Tabla 15: Endoscopio 2: cavidad abdominal

### B.4. Endoscopio 4

Nombre original: 6080HD\_flexible\*

Resolución de origen: 768x576

Resolución de trabajo: 640x360(lr)

Tipo de endoscopia: gastroscopia.

Descripción: Recorrido del esófago y exploración del estómago. Escenas de baja calidad con mucho ruido en la imagen y abundancia de reflejos.

En la Tabla 16 se listan las secuencias correspondientes al endoscopio 4.

Bag	Área mostrada	Movimiento	Duración
FLEXseq1.bag	Esófago y estómago	Avance y bamboleo	247.0 s
FLEXseq2.bag	Estómago	Bamboleo	42.9 s
FLEXseq3.bag	Estómago	Bamboleo	41.6 s
FLEXseq4.bag	Estómago	Bamboleo	80.9 s

Tabla 16: Endoscopio 4: Aparato digestivo, gastroscopia

## **C. Datos experimentales**

El objeto del Anexo C del Trabajo de Fin de Grado es el de completar la sección 5 del trabajo con datos adicionales extraídos durante los experimentos, proporcionando evidencia estadística de las conclusiones.

Se presentan Tablas similares a las presentadas en la sección 5, acompañadas de gráficos Whisker que muestran la distribución de los valores.

### **C.1. Cavidad abdominal: hígado y bazo**

En este primer apartado únicamente se adjuntan los gráficos, Figura 11, las tablas asociadas pueden encontrarse en la sección 5 con los identificadores Tabla 5, 6 y 7.

### **C.2. Cavidad abdominal: hígado y vesícula biliar**

Se presentan valores estadísticos análogos a los anteriores en las Tablas 17, 18 y 19; y su representación gráfica en la Figura 12.

### **C.3. Cavidad abdominal: intestino**

Para las secuencias del intestino procedentes del endoscopio 1 no se presentan valores estadísticos, puesto que como ya se ha explicado en la sección 5.3, el funcionamiento no es lo bastante prolongado en el tiempo para extraer valores estadísticos fiables.

En cuanto a las secuencias procedentes del segundo endoscopio, siendo el funcionamiento en estas significativamente mejor, se presentan las Tablas 20, 21 y 22, con información análoga a la del apartado anterior, así como gráficos representativos en la Figura 13.

### **C.4. Endoscopia digestiva alta: esófago y estómago**

Nuevamente en esta sección no se logra un funcionamiento lo bastante extendido en el tiempo para presentar valores estadísticos representativos.

### **C.5. Bucles en la cavidad abdominal**

Se presentan valores estadísticos análogos a los anteriores en las Tablas 23, 24 y 25; y su representación gráfica en la Figura 14.



Estado del punto del mapa	Media	Mediana	% Predicho	Rango
Predicho en la imagen	349	321	-	196 - 547
Emparejado con un ORB	116	116	36	58 - 171
No emparejado: sin un ORB cercano	195	168	53	84 - 352
No emparejado: sin un ORB similar	1	0	0	0 - 1
No cumple la condición de rigidez	38	36	11	7 - 76

Tabla 17: Tracking en hígado y vesícula. El porcentaje está referido al total de puntos predichos en la imagen.

	Media	Mediana	% Epipolar	Rango
ORBs disponibles para emparejar	674	674	-	610-736
Cumplen geometría epipolar	150	149	-	48-266
Descartado por profundidad negativa	1	1	1	0-6
Descartado por paralaje insuficiente	41	19	16	0-164
Descartado por error de reproyección	20	19	16	6-40
Descartado por escala inconsistente	3	1	1	0-12
Descartado por no cumplir rigidez	47	43	37	8-107
Puntos creados	38	34	29	11-79

Tabla 18: Gestión del mapa en hígado y vesícula: creación de puntos. El porcentaje está referido al total de puntos que cumplen con la geometría epipolar.

Estado del punto del mapa	Media	Mediana	% Total	Rango
Eliminados por no ser emparejables	16	13	25	0-37
Eliminados por no encontrarse en keyframes	7	6	12	0-20
Eliminados por no cumplir rigidez	34	33	63	3-70

Tabla 19: Gestión del mapa en hígado y vesícula: borrado de puntos. El porcentaje está referido respecto del total de puntos descartados: los no emparejables, mas los que no se encuentran en keyframes, mas los que no cumplen rigidez

Estado del punto del mapa	Media	Mediana	% Predicho	Rango
Predicho en la imagen	403	416	-	230 - 558
Emparejado con un ORB	167	171	41	93 - 242
No emparejado: sin un ORB cercano	204	217	52	75 - 293
No emparejado: sin un ORB similar	1	0	0	0 - 1
No cumple la condición de rigidez	32	28	7	12 - 59

Tabla 20: Tracking en intestino, endoscopio 3

	Media	Mediana	% Epipolar	Rango
ORBs disponibles para emparejar	636	640	-	554-708
Cumplen geometría epipolar	89	86	-	21-176
Descartado por profundidad negativa	14	12	17	1-35
Descartado por paralaje insuficiente	11	3	4	0-67
Descartado por error de reproyección	8	7	10	1-18
Descartado por escala inconsistente	1	0	0	0-3
Descartado por no cumplir rigidez	24	24	34	6-59
Puntos creados	31	25	35	6-80

Tabla 21: Gestión del mapa en intestino, endoscopio 3: creación de puntos

Estado del punto del mapa	Media	Mediana	% Total	Rango
Eliminados por no ser emparejables	12	10	26	1-30
Eliminados por no encontrarse en keyframes	3	2	5	0-7
Eliminados por no cumplir rigidez	29	26	69	9-55

Tabla 22: Gestión del mapa en intestino, endoscopio 3: borrado de puntos

Estado del punto del mapa	Media	Mediana	% Predicho	Rango
Predicho en la imagen	363	355	-	255 - 491
Emparejado con un ORB	123	121	34	74 - 174
No emparejado: sin un ORB cercano	194	191	54	109 - 296
No emparejado: sin un ORB similar	1	0	0	0 - 1
No cumple la condición de rigidez	46	44	12	19 - 78

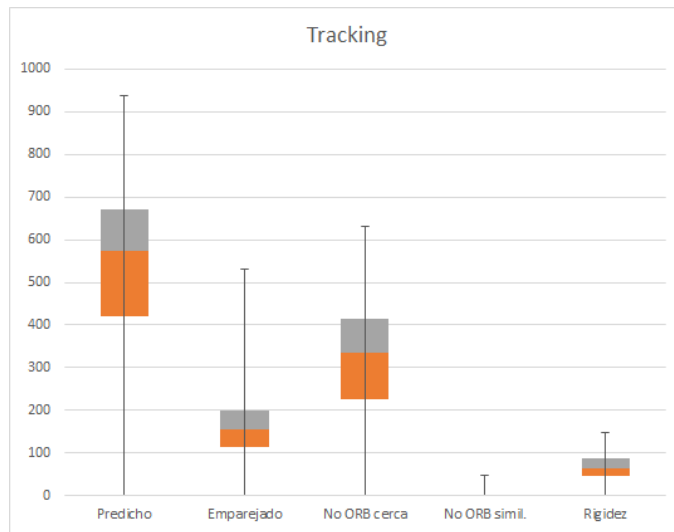
Tabla 23: Tracking en cavidad abdominal, bucle

	Media	Mediana	% Epipolar	Rango
ORBs disponibles para emparejar	665	665	-	609-715
Cumplen geometría epipolar	140	133	-	42-283
Descartado por profundidad negativa	4	3	2	0-11
Descartado por paralaje insuficiente	33	20	15	0-132
Descartado por error de reproyección	32	28	22	8-79
Descartado por escala inconsistente	3	1	1	0-10
Descartado por no cumplir rigidez	57	52	39	22-108
Puntos creados	33	29	22	9-73

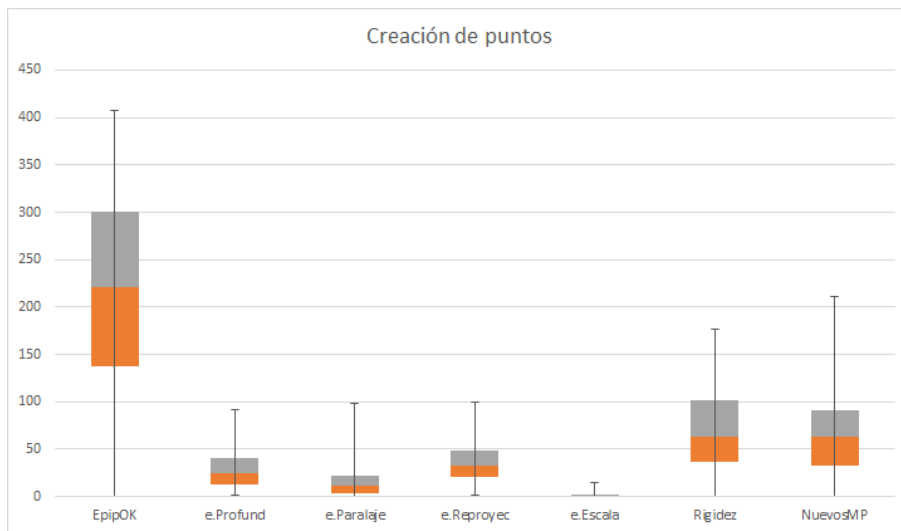
Tabla 24: Gestión del mapa en cavidad abdominal, bucle: creación de puntos

Estado del punto del mapa	Media	Mediana	% Total	Rango
Eliminados por no ser emparejables	14	13	20	0-35
Eliminados por no encontrarse en keyframes	8	6	10	0-26
Eliminados por no cumplir rigidez	44	43	70	21-78

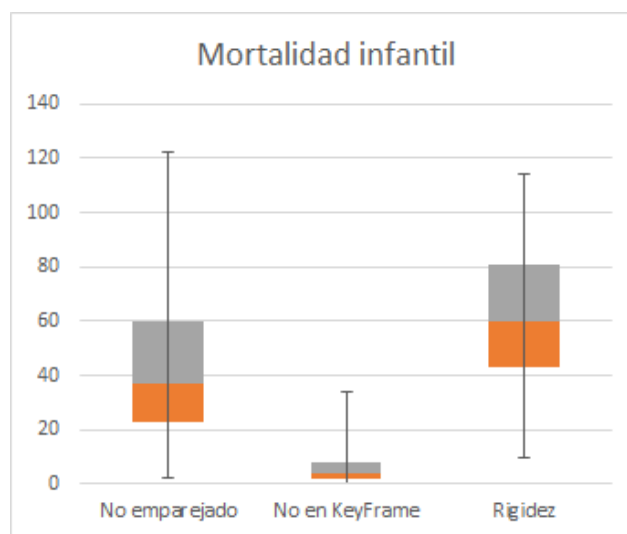
Tabla 25: Gestión del mapa en cavidad abdominal, bucle: borrado de puntos



(a)

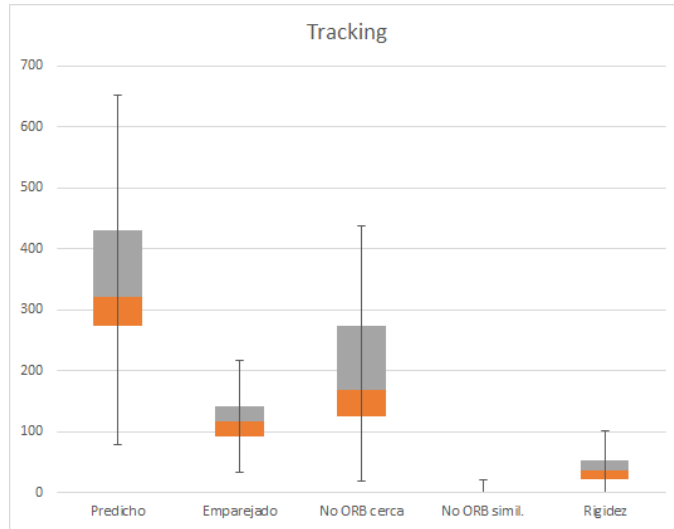


(b)

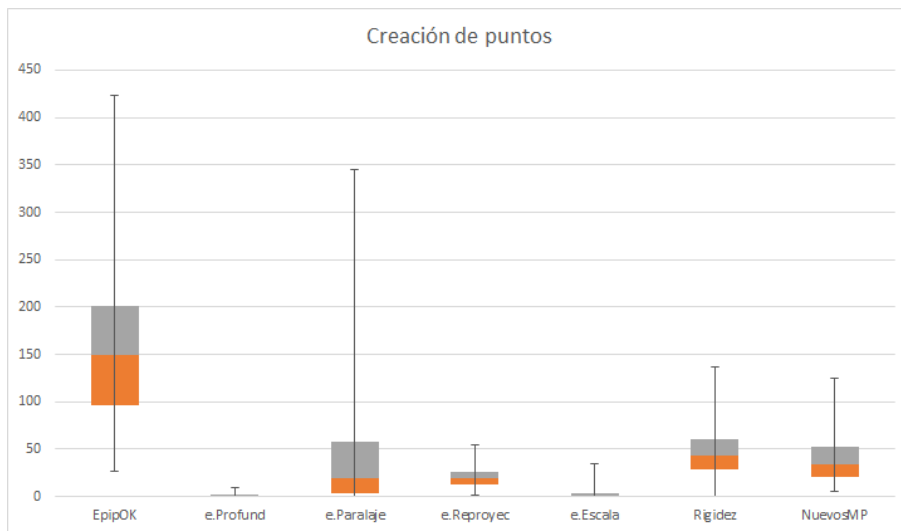


(c)

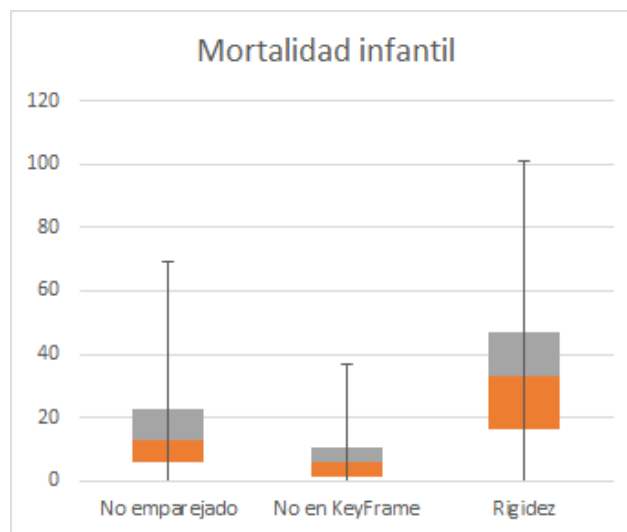
Figura 11: Whisker: cavidad abdominal, hígado y bazo



(a)

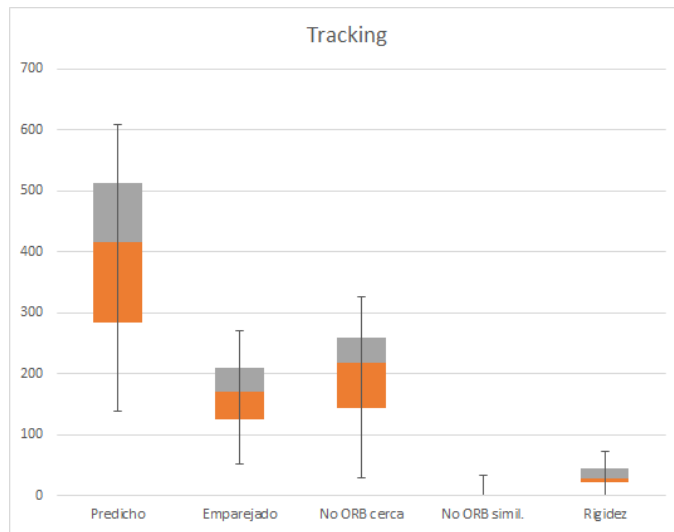


(b)

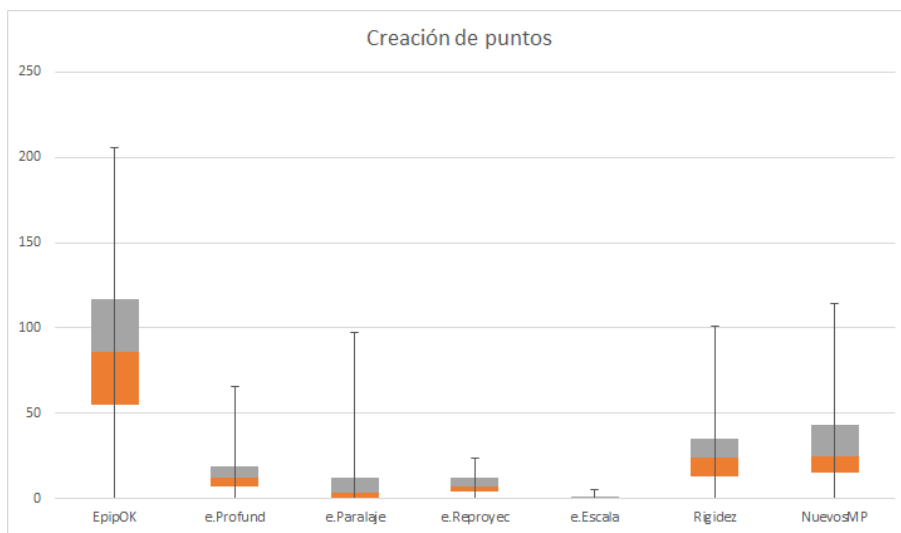


(c)

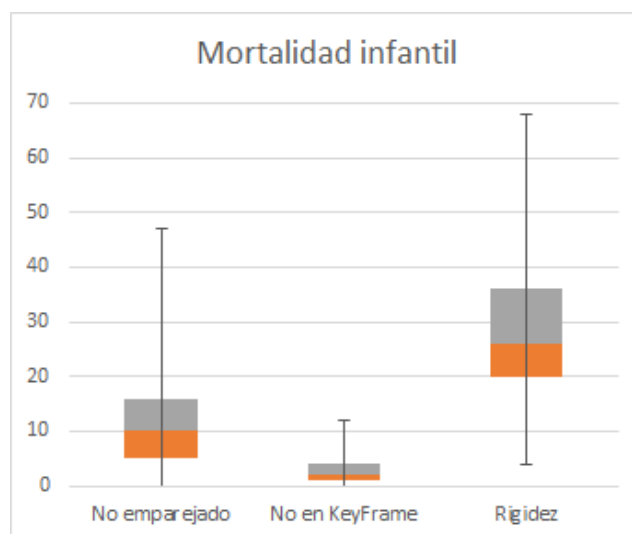
Figura 12: Whisker: cavidad abdominal, hígado y vesícula



(a)

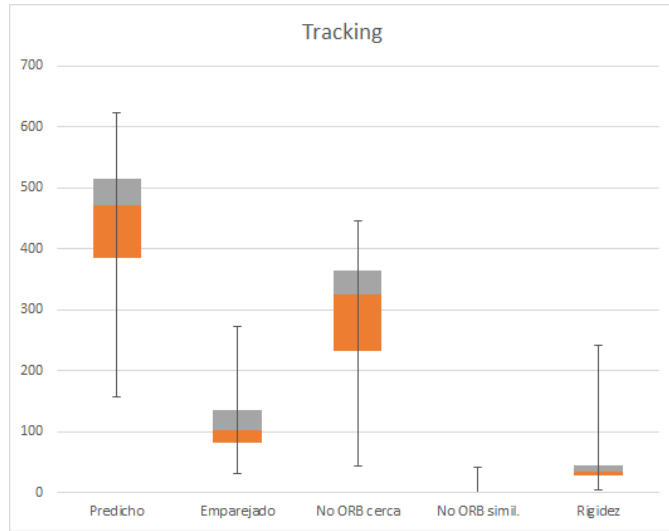


(b)

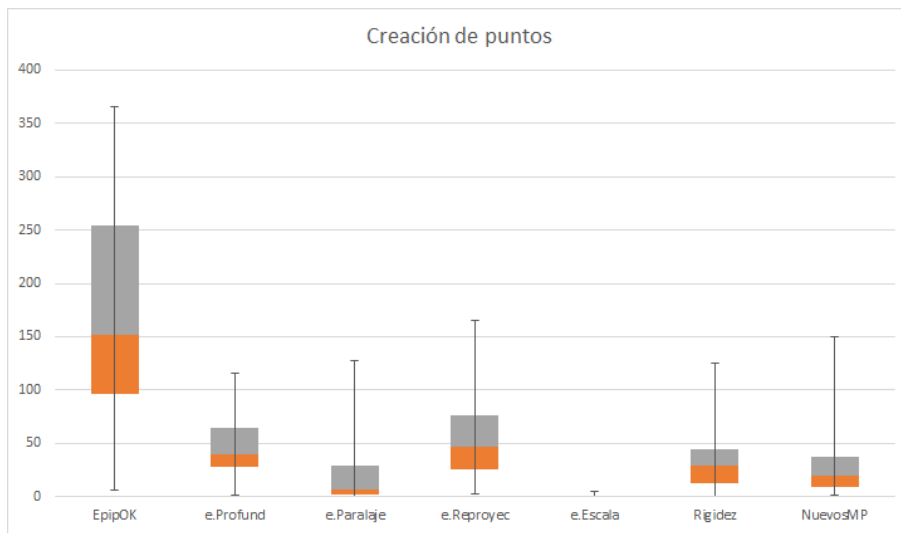


(c)

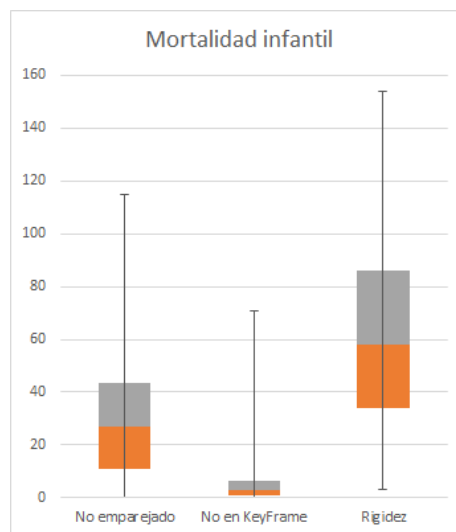
Figura 13: Whisker: cavidad abdominal, intestino, endoscopio 3



(a)



(b)



(c)

Figura 14: Whisker: cavidad abdominal, bucle