UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

DESARROLLO DE UN ALGORITMO DE STITCHING PARA SECUENCIAS DE IMÁGENES CON AMPLIOS MOVIMIENTOS DE CÁMARA

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL ELÉCTRICO

SEBASTIÁN ANDRÉS GÁLVEZ ORTIZ

PROFESOR GUÍA:
RODRIGO PALMA AMESTOY

MIEMBROS DE LA COMISIÓN:
PABLO GUERRERO PÉREZ
ANDRÉS CABA RUTTE

Este trabajo ha sido parcialmente financiado por Woodtech S.A. y Red To Green S.A.

SANTIAGO DE CHILE
2017

RESUMEN DE MEMORIA INGENIERO CIVIL ELÉCTRICO
POR: SEBASTIÁN ANDRÉS GÁLVEZ ORTIZ
FECHA: 2017
PROF. GUÍA: SR. RODRIGO PALMA AMESTOY

DESARROLLO DE UN ALGORITMO DE STITCHING PARA SECUENCIAS DE IMÁGENES CON AMPLIOS MOVIMIENTOS DE CÁMARA

En la actualidad, se han desarrollado los algoritmos de *stitching* para sintetizar el contenido de múltiples imágenes. Dependiendo del tipo de movimiento descrito por la cámara, existen diversas formas de abordar este problema, ya sea generando una imagen plana o una representación tridimensional de la escena. En particular, secuencias de imágenes capturadas con cambios grandes de la posición de la cámara, presentan un desafío para su aplicación directa. Así, el presente trabajo desarrolla una propuesta de algoritmo que incorpora técnicas del estado del arte para abordar este tipo de secuencias.

La propuesta se basa en establecer correspondencias mediante la extracción y calce de descriptores visuales y es implementada en dos fases. En la primera, se explora el uso de transformaciones de homografía para relacionar imágenes, proyectando el contenido a una imagen de referencia. En la segunda fase, se estiman simultáneamente las poses de la cámara y una representación aproximada de la escena, correspondiente a una superficie tridimensional, sobre la que se proyecta el contenido de cada vista. Para evaluar el desarrollo, se definen pruebas que incluyen la medición del error de reproyección promedio y la evaluación visual de las composiciones finales.

En los resultados obtenidos para el primer enfoque, se miden desplazamientos promedio de más de $4[px]$ al proyectar sucesivamente puntos correspondientes hacia la vista de referencia. Luego, en la composición final, se observan muchos sectores mal alineados, ya que las transformaciones obtenidas son válidas en zonas muy acotadas de la imagen, producto de las diferencias de profundidad. Estos resultados llevan a implementar la segunda fase, donde se obtienen reconstrucciones parciales de la escena con un error de reproyección promedio de $0{,}005[px]$, con desviación de $0{,}0018[px]$. Así, a pesar de la baja densidad de la nube de puntos, existe una mejora en la apreciación visual del alineamiento en la composición, además de introducir la ventaja de utilizar puntos de vista arbitrarios.

Con el trabajo realizado, se logran identificar las limitaciones del registro por homografías sobre las secuencias tratadas. Además, se presenta una propuesta que mejora la alineación del primer enfoque, al lograr combinar las distintas vistas de una secuencia sobre una representación tridimensional aproximada de la escena. Finalmente, se concluye que la propuesta desarrollada otorga una base para solucionar el problema planteado y permite identificar líneas de trabajo futuro, entre las cuales se destaca el buscar obtener una reconstrucción métrica densa de la escena que sintetice la información de todas las vistas en el modelo tridimensional.

Dedicado a todos quienes creen en mí, sobretodo a mis padres, Tito y Margarita.

Agradecimientos

En primer lugar, quiero agradecer a la empresa Woodtech S.A. y a Red To Green S.A., por abrirme sus puertas para desarrollar este trabajo y por el financiamiento otorgado. En particular, destacar la valiosa ayuda que me ofrecieron constantemente Rodrigo Palma, mi profesor guía, y todo el equipo del área de desarrollo de Woodtech S.A..

En segundo lugar, debo agradecer a mi familia por ser el pilar fundamental que me soporta constantemente, no tan sólo en la carrera, si no que en todos los aspectos de la vida.

Por último, pero no menos importante, quisiera dar una mención especial de agradecimiento a mis hermanos de la vida: Dani Zúñiga y Willy Valenzuela, quienes estuvieron constantemente a mi lado a lo largo de este arduo proceso, dándome ánimos y ayudando a distraerme cuando era necesario. También agradecer a todos los de la Rama de Tenis de Mesa de Injeniería, en especial al Feña Jorquera y a la Blanca Durán, por compartir tantos buenos momentos en torno al deporte que me apasiona, momentos sin los cuales no sería quien soy ahora.

Tabla de Contenido

1. Introducción **1**

1.1. Motivación 1

1.2. Objetivo general 2

1.3. Objetivos específicos 2

1.4. Alcances 2

1.5. Contexto general del trabajo 3

1.5.1. La empresa Woodtech S.A. 3

1.5.2. El producto *Logmeter* 3

1.6. Estructura del documento 5

2. Estado del arte **6**

2.1. Visión por computadora 6

2.2. Algoritmos de stitching 7

2.2.1. Adquisición de imágenes 9

2.2.2. Extracción de características 12

2.2.3. Registro de imágenes 14

2.2.4. Composición de imágenes 20

2.3. Herramientas 21

3. Metodología **22**

3.1. Método 22

3.1.1. Pre-procesamiento . . . 22
3.1.2. Extracción de características . . . 24
3.1.3. Establecimiento de correspondencias . . . 26
3.1.4. Metodología para el alineamiento y fusión de imágenes . . . 28
3.1.5. Algoritmo de stitching por homografías globales . . . 28
3.1.6. Algoritmo de stitching por reconstrucción de escena . . . 29
3.1.7. Reconstrucción 3D de escena . . . 31
3.1.8. Adaptación para secuencias de vehículo en movimiento . . . 35
3.1.9. Métricas . . . 37
3.2. Plan de trabajo . . . 38
3.2.1. Selección de bases de datos . . . 38

4. Resultados y Discusión **40**

4.1. Bases de datos . . . 40
4.2. Definición de pruebas . . . 43
4.3. Resultados y discusión . . . 44
4.3.1. Adaptación para secuencias de vehículo en movimiento . . . 44
4.3.2. Evaluación de la mejora de contraste sobre la extracción de características y cantidad de correspondencias establecidas . . . 46
4.3.3. Prototipo de algoritmo de stitching por homografías globales . . . 48
4.3.4. Algoritmo de stitching por reconstrucción de escena . . . 55
4.3.5. Discusión general del trabajo realizado . . . 63

5. Conclusión **65**

5.1. Trabajo futuro . . . 66

Bibliografía **67**

6. Anexos **73**

6.1. Bases de datos 73

6.2. Efecto de mejora de contraste CLAHE 82

6.3. Resultados adicionales del prototipo de algoritmo de stitching por homografías 83

6.4. Resultados adicionales del algoritmo de stitching por reconstrucción de escenas 86

Índice de Ilustraciones

Figura 1.1. Imágenes caputardas por *Logmeter* 4

Figura 2.1. Creación de panorama tolerante a paralaje. 8

Figura 2.2. Etapas de algoritmos de stitching basados en *features*. 9

Figura 2.3. Movimientos de cámara, correspondientes a rotación pura y traslación pura. [14] 10

Figura 2.4. Mejora de contraste en aplicación de análisis de retina. A la izquierda, la imagen original y, a la derecha, la imagen mejorada mediante CLAHE. [17] 10

Figura 2.5. Segmentación de persona en movimiento, usando *frame difference*. A la izquierda, una de las imágenes de entrada. A la derecha, la máscara que identifica la persona en movimiento. [21] 11

Figura 2.6. Correspondencias entre puntos de interés SIFT [25]. 12

Figura 2.7. Transformación de homografía [42]. 15

Figura 2.8. Ejemplo de resultado obtenido al aplicar transformaciones de homografía globales para la alineación de imágenes [43]. 16

Figura 2.9. Ejemplo de resultado obtenido al aplicar transformaciones de homografías duales representando un plano lejano y uno correspondiente al suelo. [44]. 16

Figura 2.10. Transformación Euclidiana [46]. 17

Figura 2.11. Ejemplo de resultado obtenido al aplicar transformaciones euclidianas de rotación para la alineación de imágenes [43]. 18

Figura 3.1. Diagrama global del algoritmo de stitching propuesto. 23

Figura 3.2. Histograma y curva de función de distribución acumulativa antes y después de la ecualización del histograma. 24

Figura 3.3. Limite de corte para acotar amplificación de contraste. 24

Figura 3.4. Comparación de filtro de difusión no lineal, en la parte superior, y filtro gaussiano, en la inferior, para niveles de escala equivalentes. 25

Figura 3.5. Cálculo de descriptor *Modified-Local Difference Binary* (M-LDB), para un nivel de escala. 26

Figura 3.6. Geometría de modelo de cámara *pinhole* [13]. 30

Figura 3.7. Proyección de vistas desde distintas poses de cámara, sobre superficie aproximada a partir de nube de puntos reconstruida. 32

Figura 3.8. Diagrama de flujo de la reconstrucción de escena a partir de la estimación del conjunto de cámaras y los puntos de interés calzados. 32

Figura 3.9. Reproyección de puntos de la escena 34

Figura 3.10. Modelo de superficie NURBS con puntos de control que definen su forma. [74] 34

Figura 3.11. Diagrama global del algoritmo de stitching adaptado a presencia de objetos en movimiento. 36

Figura 3.12. Carta Gantt de la planificación del trabajo. 38

Figura 4.1. Ejemplo de imágenes de la secuencia 'Temple'. 41

Figura 4.2. Ejemplo de imágenes de la secuencia 'Fountain'. 41

Figura 4.3. Ejemplo de imágenes de la secuencia 'HerzJesu'. 41

Figura 4.4. Ejemplo de imágenes de la secuencia 'SceauxCastle'. 41

Figura 4.5. Secuencia de imágenes 'PaisajeNavarino'. 42

Figura 4.6. Ejemplo de imágenes de la secuencia 'Truck1R'. 42

Figura 4.7. Ejemplo de imágenes de la secuencia 'Truck1L'. 42

Figura 4.8. Ejemplo de imágenes de la secuencia 'Truck2R'. 43

Figura 4.9. Ejemplo de imágenes de la secuencia 'Truck2L'. 43

Figura 4.10. Pares de imágenes de referencia, para ejemplificar segmentación y filtrado de calces de fondo en secuencia 'Truck1R'. 44

Figura 4.11. Ejemplo de segmentación por movimiento para la primera imagen de la secuencia 'Truck1R'. 45

Figura 4.12. Ejemplo de segmentación por movimiento para la penúltima imagen de la secuencia 'Truck1R'. 45

Figura 4.13. Ejemplo de filtrado de calces de fondo por desplazamiento mínimo en el primer par de imágenes de la secuencia 'Truck1R'. 45

Figura 4.14. Ejemplo de filtrado de calces de fondo por desplazamiento mínimo en el último par de imágenes de la secuencia 'Truck1R'. 45

Figura 4.15. Efecto de mejora de contraste CLAHE sobre total de correspondencias validadas por secuencia. 47

Figura 4.16. Prueba inicial del prototipo de stitching por homografías, sobre secuencia 'Paisaje Navarino'. 49

Figura 4.17. Desplazamiento promedio en pares de vistas consecutivas en secuencia 'Fountain'. El desplazamiento acumulado para llevar la última vista hacia la primera es de 4,14$[px]$ 50

Figura 4.18. Desplazamiento promedio en pares de vistas consecutivas en secuencia 'Truck2L'. El desplazamiento acumulado para llevar la última vista hacia la primera es de 6,81$[px]$ 50

Figura 4.19. Resultado de la composición de la secuencia que abarca el intervalo $[3, 9]$ de 'Truck2L', mediante registro por homografías globales hacia el plano de referencia de la primera imagen. 51

Figura 4.20. Resultado de la composición de la secuencia 'Fountain', mediante registro por homografías globales hacia primera imagen. 52

Figura 4.21. Resultado de stitching de la secuencia que abarca el intervalo $[3, 9]$ de 'Truck2L', mediante registro por homografías globales hacia primera imagen. 52

Figura 4.22. Resultado de stitching de la secuencia 'Fountain', mediante registro por homografías globales hacia primera imagen. 53

Figura 4.23. Resultado de la reconstrucción de escena sobre la secuencia 'Temple'. 56

Figura 4.24. Resultado de la reconstrucción de escena sobre la secuencia 'HerzJesu'. 57

Figura 4.25. Resultado de la reconstrucción de escena sobre la secuencia 'Fountain'. 58

Figura 4.26. Resultado en dos vistas de la composición sobre modelo 3D, correspondiente a la secuencia 'Fountain'. 59

Figura 4.27. Resultado en dos vistas de proyecciones de imágenes de la secuencia sobre modelo 3D, correspondiente a la secuencia 'Fountain'. 60

Figura 4.28. Resultado en dos vistas de la composición sobre modelo 3D, correspondiente a la secuencia 'Truck2L'. 61

Figura 4.29. Resultados parciales de la composición al proyectar distintas imágenes de la secuencia 'Truck2L' sobre modelo 3D. 62

Figura 6.1. Secuencia de imágenes 'Temple'. 75

Figura 6.2. Secuencia de imágenes 'Fountain'. 76

Figura 6.3. Secuencia de imágenes 'PaisajeNavarino'. 76

Figura 6.4. Secuencia de imágenes 'HerzJesu'. 77

Figura 6.5. Secuencia de imágenes 'Truck1R'. 78

Figura 6.6. Secuencia de imágenes 'Truck1L'. 79

Figura 6.7. Secuencia de imágenes 'Truck2R'. 80

Figura 6.8. Secuencia de imágenes 'Truck2L'. 81

Figura 6.9. Desplazamiento promedio en pares de vistas consecutivas en secuencia 'Temple'. El desplazamiento acumulado para llevar la última vista hacia la primera es de 4,68$[px]$ 83

Figura 6.10. Desplazamiento promedio en pares de vistas consecutivas en secuencia 'Truck1R'. El desplazamiento acumulado para llevar la última vista hacia la primera es de 7,68$[px]$ 84

Figura 6.11. Resultado de la composición de la secuencia 'Temple', mediante registro por homografías globales hacia el plano de referencia de la primera imagen. . 84

Figura 6.12. Resultado de la composición de la secuencia 'Truck1R', mediante registro por homografías globales hacia el plano de referencia de la primera imagen. 85

Figura 6.13. Resultado de la composición de la secuencia 'Truck2L' completa, mediante registro por homografías globales hacia el plano de referencia de la primera imagen. 85

Figura 6.14. Resultado de la reconstrucción de escena sobre la secuencia 'Temple'. 86

Figura 6.15. Resultado de la reconstrucción de escena sobre la secuencia 'Sceaux-Castle'. 87

Figura 6.16. Resultado de la reconstrucción de escena sobre la secuencia 'HerzJesu'. 88

Figura 6.17. Resultado de la reconstrucción de escena sobre la secuencia 'Truck1R'. 89

Figura 6.18. Resultado de la reconstrucción de escena sobre la secuencia 'Truck1L'. 90

Figura 6.19. Resultado de la reconstrucción de escena sobre la secuencia 'Truck2R'. 91

Índice de Tablas

Tabla 4.1. Aumento porcentual promedio de cantidad de puntos de interés detectados para las distintas secuencias. 47

Tabla 4.2. Número de puntos reconstruidos y errores de reproyección promedio inicial y final para las secuencias evaluadas. 55

Tabla 6.1. Resumen de bases de datos escogidas, con detalle de la resolución de las imágenes utilizadas y los parámetros intrínsecos de las cámaras, además de la transformación realizada respecto de las imágenes de la base de datos original de donde se obtuvieron las secuencias. 74

Tabla 6.2. Efecto de mejora de contraste CLAHE sobre cantidad de puntos de interés detectados y total de correspondencias validadas por secuencia 82

Capítulo 1

Introducción

1.1. Motivación

En la actualidad, existen diversas situaciones que requieren observación de una o más imágenes digitales, como por ejemplo, la búsqueda en mapas satelitales, el análisis de radiografías y el control de calidad en procesos productivos. En muchos de estos casos, la observación es, en realidad, realizada sobre una síntesis creada a partir del contenido de múltiples imágenes, pues, de lo contrario, esta constituiría un proceso lento e ineficiente. Esta síntesis puede ser alcanzada gracias a técnicas de las áreas de procesamiento de imágenes y de visión por computadora, las que se encuentran en un nivel de desarrollo tal, que es posible generar vistas panorámicas en 360° combinando múltiples imágenes bajo variadas condiciones, o también reconstruir modelos tridimensionales de las escenas capturadas, incluso desde un *smartphone*.

Por otro lado, en el contexto de la industria forestal, existen múltiples necesidades relacionadas con el control de calidad de la madera. Actualmente, uno de los procesos relevantes para este propósito consiste en la inspección visual de los bancos de madera transportados por camiones. Con tal objetivo, es común utilizar operarios que observen directamente el cargamento del camión, opción que no siempre es la mejor.

La empresa Woodtech S.A., dedicada a desarrollar soluciones tecnológicas para distintas problemáticas de la industria forestal, posee un producto capaz de reconstruir, mediante mediciones realizadas con láser, el modelo tridimensional de los camiones con su cargamento. Este producto, incluye una herramienta de inspección visual, que despliega múltiples fotografías del cargamento de madera a un operario, quien debe revisarlas una a una. Si bien esta solución ayuda al proceso de observación de la carga, el hecho de tener que recorrer múltiples imágenes no relacionadas entre sí, puede generar confusión en el operario, haciendo el proceso lento y poco eficiente.

Con estos precedentes, Woodtech S.A. ve una oportunidad para mejorar la herramienta de inspección visual de su producto, por lo que desea investigar la factibilidad técnica de utilizar algoritmos de stitching sobre las imágenes capturadas, con el objetivo de simplificar el proceso de observación, sintetizando la información desplegada. Por otro lado, en una línea paralela

de negocio, la empresa también desea estudiar la capacidad de utilizar la información presente en estas imágenes para complementar las mediciones con láser realizadas a los camiones.

Las imágenes capturadas por el producto descrito, presentan un desafío para la aplicación de los algoritmos de stitching, ya que se producen cambios grandes de traslación entre la captura de cada cuadro de la secuencia. Por este motivo, el presente trabajo surge para dar solución a la problemática general de utilizar algoritmos de stitching sobre secuencias de imágenes con movimientos amplios de la cámara, lo que, a su vez, solucionaría el problema planteado por la empresa. Al mismo tiempo, considerando el estado actual de estos algoritmos, permite mostrar el potencial de realizar la reconstrucción tridimensional a partir de estas imágenes.

1.2. Objetivo general

El objetivo principal de este trabajo es desarrollar un algoritmo de stitching para sintetizar la información desplegada en secuencias de imágenes con importantes traslaciones de la cámara, tomando en consideración los intereses de la empresa Woodtech S.A. de aplicarlo sobre las imágenes de camiones con cargamento de madera y, a la vez, explorar su uso potencial para la reconstrucción tridimensional del camión.

1.3. Objetivos específicos

Dentro de los objetivos específicos planteados se encuentran:

- Comprender las etapas involucradas en los algoritmos de stitching y su relación con la reconstrucción de escenas a partir de imágenes, a través de la revisión bibliográfica.
- Determinar las etapas adicionales necesarias para aplicar estos algoritmos sobre las imágenes de la herramienta de inspección visual.
- Proponer un diseño del algoritmo de stitching para fusionar imágenes capturadas desde una cámara con importantes movimientos de traslación.
- Implementar el algoritmo mediante bloques de software capaces de realizar las etapas del diseño propuesto.
- Evaluar y analizar los resultados del desarrollo realizado mediante una métrica previamente definida.

1.4. Alcances

El presente trabajo se plantea como el desarrollo de una propuesta de algoritmo de stitching, la cual requiere de etapas posteriores de desarrollo para su incorporación en algún

producto comercial. En este sentido, los alcances del algoritmo abarcan la fusión de secuencias ordenadas de imágenes, considerando que se tiene disponibilidad de toda la secuencia al momento de aplicar el algoritmo. Las secuencias tratadas en este desarrollo poseen cambios relativamente grandes en la pose de la cámara que las captura, por lo que el alcance de la propuesta permitiría la fusión de secuencias en las que estos cambios poseen un contenido compartido suficiente para establecer una relación entre las imágenes consecutivas de la secuencia.

En la aplicación de este algoritmo al producto *Logmeter*, se consideran secuencias capturadas desde una cámara estática con vista lateral de un camión en movimiento, lo que es un paso fundamental para lograr, eventualmente, unificar todas las vistas disponibles. La propuesta se adapta a este tipo de captura y presenta una solución para la fusión del contenido de las imágenes de estas secuencias. Esta solución, otorga una implementación base, que tiene como foco central el alineamiento de imágenes, por lo que requiere etapas posteriores de desarrollo llegar a una versión comercial.

1.5. Contexto general del trabajo

Debido a que este proyecto está motivado por una necesidad específica de la empresa Woodtech S.A., es necesario introducir algunos aspectos relevantes de esta compañía y del producto en el que se busca aplicar este trabajo. A continuación, se presenta el área de la industria en la que se desenvuelve la empresa y se describen estos aspectos.

1.5.1. La empresa Woodtech S.A.

La empresa Woodtech S.A., una compañía de Red to Green S.A. [1], se dedica a desarrollar soluciones tecnológicas para diversas necesidades en procesos industriales, especialmente aquellas relacionadas con problemas de medición en la industria forestal. En este contexto, una de las líneas de productos principales es la de modelamiento e inspección de troncos [2].

El desarrollo de este proyecto se realiza en colaboración directa con la empresa, que facilita un puesto de trabajo en sus oficinas y otorga el acceso a imágenes capturadas con el producto involucrado. Además, se cuenta con el apoyo técnico del equipo de desarrollo para complementar el trabajo realizado y facilitar su integración.

1.5.2. El producto *Logmeter*

El *Logmeter* es un producto diseñado para automatizar una serie de procesos de medición de los troncos que llegan a una planta y destaca por realizar la estimación automática del volumen y otros parámetros del banco de madera de un camión que atraviesa un portal con múltiples sensores láser [3]. Adicionalmente, mediante cámaras estáticas, captura una serie de fotografías del vehículo y su carga.

Este producto, además de realizar una reconstrucción tridimensional del camión mediante escaneo láser, incluye una herramienta de inspección visual, la cual presenta una interfaz gráfica que despliega una a una las imágenes capturadas, permitiendo al operario realizar la inspección visual de la carga de madera. Es en esta herramienta donde se identifica la oportunidad de mejora que motiva el presente trabajo, ya que las imágenes son mostradas de manera individual generando ciertos problemas para su uso. Por ejemplo, para observar la carga completa, se deben recorrer muchas imágenes por cada vista del camión, las cuales, además, son redundantes. Estas condiciones pueden hacer que el proceso de inspección se vuelva tedioso e ineficiente, por lo que la síntesis de la información desplegada permitiría mejorar esta situación.

En una línea de negocio paralela, la empresa muestra interés por complementar el modelo tridimensional reconstruido por láser, usando la información de las imágenes y, eventualmente, poder sustituir los sensores láser por cámaras para crear un nuevo producto.

En la Fig. 1.1 se muestran algunas de las imágenes desplegadas en el sistema, capturadas por una cámara con vista lateral al camión. Esta forma de adquisición conlleva múltiples desafíos para el diseño de un algoritmo de stitching, entre los cuales se destacan:

Figura 1.1: Imágenes del cargamento de troncos en un camión, capturadas por una cámara con vista lateral a este. [4]

- **Imágenes de fondo fijo y vehículo de carga en movimiento:** La fusión de imágenes suele ser realizada sobre escenas estáticas, por lo que supone un desafío su aplicación al caso de un vehículo de carga en movimiento.
- **Alto contenido de textura:** La corteza de los troncos posee una textura que puede dificultar la adquisición de correspondencias correctas entre imágenes.
- **Diferentes puntos de vista y en perspectiva:** Establecer relaciones entre las vistas de la escena que presentan grandes cambios de perspectiva conlleva el uso de modelos complejos que incorporen estos cambios.

1.6. Estructura del documento

Lo que resta del documento sigue la estructura descrita a continuación.

El Capítulo 2 presenta una detallada revisión bibliográfica de los algoritmos de stitching y de las distintas opciones existentes en la literatura para realizar cada etapa, abordando principalmente aquellas que corresponden al estado del arte, donde se presenta la relación existente con los algoritmos de reconstrucción de escena.

El Capítulo 3 aborda la metodología del trabajo otorgando una concepción general del diseño propuesto y las fases generales del desarrollo. Luego, se profundiza cada parte del mismo describiendo en detalle las técnicas más relevantes del algoritmo. Finalmente, se presenta el plan de trabajo para el desarrollo, definiendo los criterios para escoger datos de prueba y las métricas que serán utilizadas para evaluar las distintas fases del desarrollo.

El Capítulo 4 presenta las bases de datos utilizadas, define las pruebas pertinentes y expone los resultados obtenidos, acompañados de la discusión sobre los mismos y los posibles puntos de mejora.

El Capítulo 5 finaliza el documento con las conclusiones relevantes sobre el proyecto y las principales opciones de trabajo futuro.

Capítulo 2

Estado del arte

En este capítulo, se introduce el marco teórico que permite al lector comprender el área específica de ingeniería en la que se desarrolla el presente trabajo. Se comienza describiendo brevemente el concepto de visión por computadora y luego, se profundiza sobre los algoritmos de stitching, describiendo las técnicas que permiten abordar los problemas involucrados. Además, se presentan trabajos correspondientes al estado del arte, que son relevantes para este proyecto.

2.1. Visión por computadora

La visión por computadora es una disciplina reciente, que busca describir el mundo observado en una o más imágenes y reconstruir sus propiedades. Para esto, los trabajos del área de visión por computadora se enfocan en modelar, replicar y aumentar las capacidades del sistema visual humano, usando hardware y software computacional, que incorpora el conocimiento adquirido por múltiples disciplinas como ciencias de la computación, ingeniería eléctrica, matemática, fisiología, biología y ciencias cognitivas.

Esta disciplina busca dar una interpretación al contenido de las imágenes, por lo que requiere, en primera instancia, identificar características de la escena a partir de la información disponible en los pixeles, mediante métodos que incluyen: detección de bordes; segmentación; extracción y descripción de puntos de interés; entre otros. En segundo lugar, se utilizan estas características para reconocer los objetos presentes en la escena y extraer sus propiedades. Es importante mencionar que el entendimiento de una escena del mundo real, puede requerir etapas como la obtención de parámetros de la cámara utilizada en cada captura, e incluso la reconstrucción de un modelo tridimensional que represente las estructuras observadas, por lo que la visión computacional no se limita al procesamiento de pixeles.

Así, entre las aplicaciones que hacen uso de la visión por computadora, se pueden encontrar: sofisticados sistemas de vigilancia; aplicaciones de realidad aumentada; vehículos autónomos; control automático de calidad en procesos industriales; modelamiento 3D de edificios; asistencia en diagnóstico con imágenes médicas; entre otros. [5]

2.2. Algoritmos de stitching

Utilizando técnicas de visión por computadora, se pueden desarrollar algoritmos encargados de combinar múltiples imágenes para obtener una imagen coherente capaz de representar información que abarca un campo visual mayor al de cada imagen por sí sola, en la medida que exista una conexión entre ellas. La técnica utilizada por estos algoritmos se conoce en la literatura como *stitching* o *mosaicing*. En el área de procesamiento digital de imágenes, ésta se desarrolló en un principio como un procedimiento capaz de generar automáticamente vistas panorámicas de escenas amplias y lejanas, combinando imágenes captadas en movimientos acotados de una única cámara [6]. Los algoritmos actuales de stitching son capaces de analizar el contenido de la escena para reconocer panoramas en un conjunto desordenado de imágenes, encontrar una forma óptima de fusionarlas y eliminar artefactos que se pueden generar en el proceso de combinación, entre otras funcionalidades. Incluso, en la actualidad es posible realizar la fusión de imágenes sobre modelos 3D de una escena, permitiendo una visualización distinta del resultado [7].

Las principales dificultades de los algoritmos clásicos de stitching se presentan en escenas que contienen alguna de las siguientes características:

- **Paralaje:** Si se considera una imagen donde se observa una escena estática con un fondo lejano y un objeto cercano, al obtener otra imagen de la misma escena con distinto punto de vista, se observará un gran cambio en la posición del objeto cercano, mientras que el fondo sufrirá cambios muy leves.
- **Objetos móviles:** Al tomar dos imágenes de una escena estática donde existe un objeto en movimiento, se dificulta obtener las correspondencias que sirven para alinear adecuadamente las imágenes, ya que la relación entre las zonas estáticas es diferente a la de las zonas en movimiento.
- **Grandes cambios de perspectiva:** Dificultan considerablemente el establecimiento de correspondencias entre imágenes, siendo un campo activo de investigación.

Una de las aplicaciones más conocidas del stitching, consiste en la función presente en las cámaras digitales que permite generar vistas panorámicas al momento de la captura. Sin embargo, las posibles aplicaciones incluyen, por ejemplo, la generación de imágenes que sirvan de entrada a sistemas de aprendizaje de máquinas para detección de productos fuera de stock en góndolas de tiendas de retail [9], generación de vistas aéreas utilizadas en la creación de mapas, vigilancia aérea o uso militar [10], entre otras. Todas estas aplicaciones dan a entender que los algoritmos de stitching pueden ser utilizados en cualquier ámbito donde se requiera fusionar imágenes, haciéndolo un tópico que motiva constantemente nuevas investigaciones. Es por ello que se ha visto últimamente un desarrollo más acabado de este campo, permitiendo incluso la generación de imágenes con vistas de 360° [11]. En la Fig. 2.1 se presenta, a modo de ejemplo, el resultado un algoritmo de stitching que logra eliminar artefactos generados por el paralaje inducido al variar el punto de vista en la captura, gracias a que incorpora información sobre el contenido de la escena al hacer uso de herramientas de visión computacional [8].

En la literatura, se suele utilizar una clasificación de los algoritmos de stitching en dos grandes clases dependiendo del método utilizado para encontrar correspondencias entre las

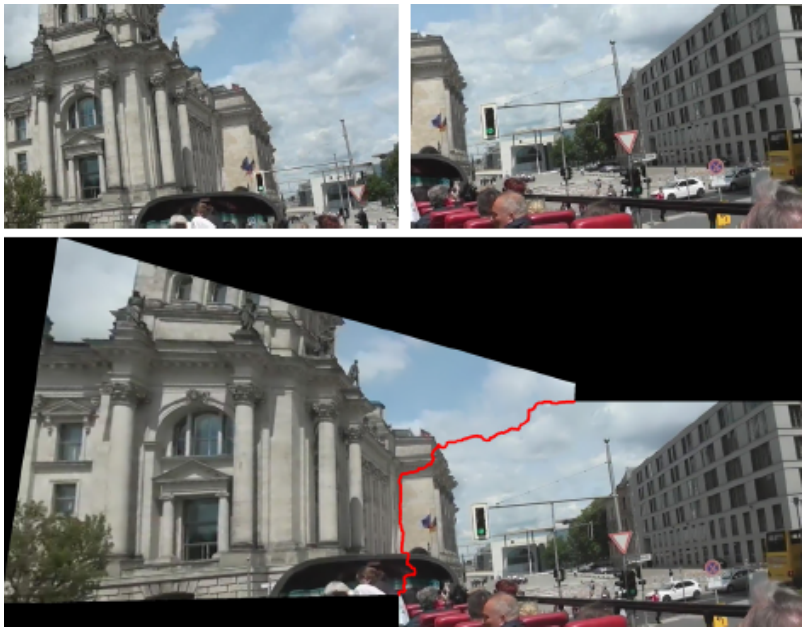


Figura 2.1: Creación de panorama tolerante a paralaje (notar cambio de la ubicación del semáforo). En la parte superior, se muestran las dos imágenes de entrada y, en la inferior, el panorama obtenido con el borde resaltado. [8]

imágenes: los *directos* y los basados en puntos de interés (*features*). El primero trabaja comparando directamente las intensidades de pixeles, siendo altamente susceptible a cambios de exposición y rotaciones, mientras que, el segundo, extrae características de ciertos puntos de interés en las imágenes, adquiriendo, por ejemplo, información invariante a intensidad, escala y rotación como la orientación y magnitud de gradientes para cada uno de estos puntos y así encontrar correspondencias de forma más robusta y en una mayor variedad de casos. En los trabajos pertenecientes al estado del arte, es ampliamente preferido el stitching basado en *features* debido al aumento de robustez y mejora en resultados. [5]

Según [12], los algoritmos de stitching basados en *features* siguen un modelo común con el flujo mostrado en la Fig. 2.2, que incluye, a grandes rasgos, las siguientes etapas.

1. **Adquisición de imágenes**: Determina la calidad de las imágenes que servirán de entrada al sistema y el tipo de movimiento realizado en la captura: traslación pura, rotación pura o combinaciones.
2. **Extracción de características**: Se identifican y describen puntos de interés que sirven para establecer correspondencias entre el contenido compartido en distintas imágenes.
3. **Registro de imágenes:** Consiste en identificar relaciones coherentes entre las distintas imágenes a partir de correspondencias encontradas entre ellas y obtener transformaciones geométricas que las representen.
4. **Composición:** Permite fusionar las imágenes en una superficie adecuada, combinando la información disponible para la obtención de una única imagen compuesta que permitiría una visualización acorde a la aplicación, evitando la generación de artefactos.

A continuación, se describe en mayor detalle cada una de las etapas y algunas de las técnicas más relevantes para realizar cada una de ellas.

Figura 2.2: Etapas de algoritmos de stitching basados en *features*.

2.2.1. Adquisición de imágenes

En cualquier sistema de visión por computadora, se debe tener especial atención en la forma de adquirir las imágenes de entrada, ya que esta etapa puede incidir fuertemente en los resultados finales del mismo. En el caso particular del stitching, además de los parámetros de exposición, enfoque y sensibilidad, es importante tener en cuenta que una escena se puede capturar con distintos movimientos de la cámara, lo que afecta el proceso de encontrar una alineación apropiada entre las imágenes. De acuerdo a esto, es importante poseer un modelo adecuado de la cámara con que se capturan las imágenes, además de las transformaciones a las que es sujeta.

Para modelar una cámara, se suele utilizar el modelo *pinhole*, que utiliza una transformación proyectiva para establecer una relación entre los puntos 3D del mundo real y los puntos 2D en el plano de la imagen, de acuerdo a la longitud focal f de la cámara. Este sencillo modelo de cámara *pinhole*, puede ser refinado para obtener resultados más realistas mediante la inclusión de parámetros que modelan la distorsión del lente, entre otras posibilidades. En su versión más simple, todos los parámetros del sensor involucrados en el modelo son expresados en una **matriz de parámetros intrínsecos** (K), mientras que, al representar la cámara como un objeto con una pose determinada, los **parámetros extrínsecos** de la cámara corresponden a una transformación euclidiana, conformada por una matriz de rotación y una traslación ($[R|t]$). Dependiendo de la información disponible, puede ser necesario estimar estos parámetros mediante un proceso conocido como **calibración**. En la Parte I de [13], se describe detalladamente la geometría proyectiva involucrada, además de los modelos *linear pushbroom* y de cámara afín.

Movimiento de cámara

En el modelo *pinhole*, la cámara es tratada como un objeto con una pose determinada, por lo que al momento de realizar las distintas capturas de una escena se pueden producir cambios de posición y orientación, los cuales corresponden a transformaciones rígidas en un

espacio euclidiano. Según los cambios producidos, es posible distinguir capturas realizadas mediante movimientos de rotación pura o traslación pura, como se muestra en la Fig. 2.3, o mediante una combinación de ambos. Cualquiera sea el caso, es importante mencionar que debe existir contenido compartido entre las imágenes para hacer posible la aplicación de un algoritmo de stitching.

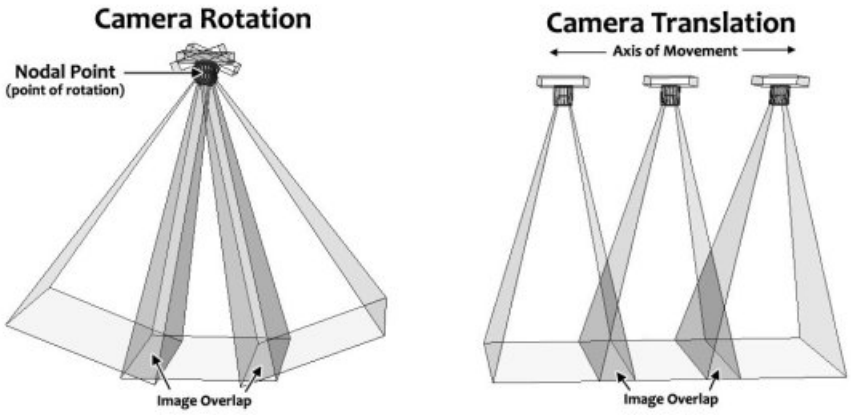


Figura 2.3: Movimientos de cámara, correspondientes a rotación pura y traslación pura. [14]

En la actualidad, existen distintas formas de medir el movimiento de la cámara de manera precisa. En [15], se presenta un esquema de calibración que aprovecha una unidad de medición inercial interna (IMU). Por otro lado, se encuentran las técnicas de odometría visual, las cuales permiten estimar los cambios de pose a partir del contenido capturado en múltiples cuadros, sin requerir hardware adicional [16].

Mejora de contraste

Independiente del modelo o el movimiento descrito por la cámara, una imagen capturada es susceptible a múltiples alteraciones respecto de la escena real que pretende representar, ya sea debido a condiciones ambientales, ruido de sensor, o a tiempos de exposición inadecuados. Por esto, es importante compensar estos efectos para obtener una imagen que aproveche la información disponible de la mejor manera posible para su uso en etapas posteriores.

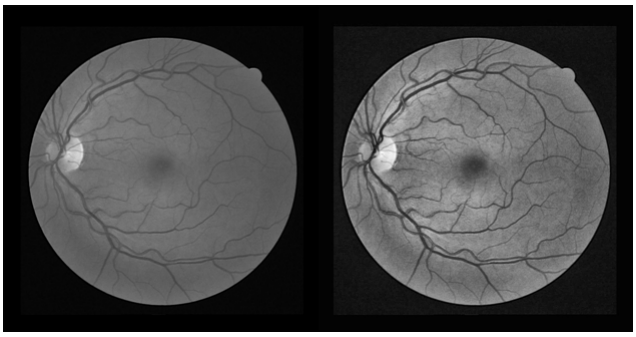


Figura 2.4: Mejora de contraste en aplicación de análisis de retina. A la izquierda, la imagen original y, a la derecha, la imagen mejorada mediante CLAHE. [17]

Un mecanismo usual para la mejora de imágenes, corresponde a la ecualización de histogramas, que busca aprovechar todo el rango de intensidades disponible, mejorando el contraste de la imagen. Si bien la ecualización puede ser realizada de manera global, existen métodos que ecualizan los histogramas de forma local, entre los que se encuentra *Contrast Limited Adaptive Histogram Equalization* (CLAHE) [18], cuyos resultados se ejemplifican en la Fig. 2.4. Este y otros métodos de ecualización de histograma se encuentran descritos en la Sección 3.1.4 de [5].

Segmentación

En ciertas aplicaciones, es necesario realizar un proceso de segmentación para aislar las zonas de interés de la imagen de las zonas que son identificadas como fondo. En [19], se presenta el método *watershed segmentation*, que utiliza el contenido morfológico de la imagen para identificar zonas separadas por contornos. Algunas formas de segmentación buscan aislar un objeto en movimiento de un fondo estático, mientras que otros, a partir de la información del color pueden identificar las zonas de la imagen que corresponden a un objeto determinado. Como se ejemplifica en la Fig. 2.5, el objetivo final de esta etapa es obtener una máscara identificando la zona de interés. En el Capítulo 5 de [5], se describe detalladamente este tópico, presentando un amplio marco teórico.

Respecto de las técnicas para segmentar objetos en movimiento, se identifican, por un lado, aquellas utilizadas comúnmente en sistemas de vigilancia y que crean un modelo del fondo, el cual es sustraído para identificar objetos de interés [20]. Por otro lado, se encuentran las que realizan *frame difference*, que identifican las zonas de mayor movimiento mediante la resta de cuadros sucesivos [21],[22]. Estos algoritmos usualmente requieren una etapa de procesamiento adicional que elimine el ruido presente e identifique zonas conexas de tamaño significativo para la aplicación. En la Sección 3.3.4 de [5] se presentan las técnicas de análisis de zonas conexas, tanto para imágenes en escala de grises como en imágenes binarias. Modelos más complejos de segmentación de objetos en movimiento son capaces de estimar la velocidad del objeto y validar que las regiones encontradas se muevan a esa velocidad [23].

Figura 2.5: Segmentación de persona en movimiento, usando *frame difference*. A la izquierda, una de las imágenes de entrada. A la derecha, la máscara que identifica la persona en movimiento. [21]

2.2.2. Extracción de características

La siguiente etapa del proceso de stitching corresponde a la extracción de características, que busca describir el contenido de la imagen, permitiendo hacer reconocimiento del mismo en otra imagen. Dependiendo de la aplicación, es deseable que las características extraídas de una imagen sean invariantes a cambios de iluminación, escala, orientación e incluso de perspectiva. Sin embargo, no todos los métodos desarrollados son capaces de lograr esto y aquellos que se acercan, aumentan considerablemente el costo computacional de la extracción.

Las características más utilizadas son extraídas en base a puntos de interés en la imagen, generalmente identificados como esquinas, y describiendo su vecindad, como se muestra en la Fig. 2.6. Por otro lado, también es posible extraer características desde bordes, líneas e incluso regiones completas, permitiendo describir directamente la forma o textura de los objetos. En la literatura, se puede encontrar una gran cantidad de alternativas para la extracción de características y, al ser un área de estudio en desarrollo, cada año se presentan nuevos detectores y descriptores. En [24], se realiza una detallada revisión de varios de los métodos existentes.

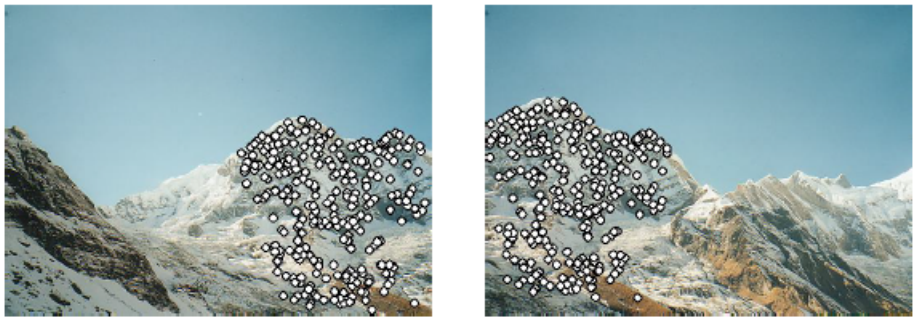


Figura 2.6: Correspondencias entre puntos de interés SIFT [25].

Para la extracción de características basada en puntos de interés, es necesario primero realizar una detección de su ubicación, para posteriormente obtener una descripción adecuada del contenido que rodea a ese punto. Entre los algoritmos que realizan únicamente detección destacan: el clásico detector de esquinas de Harris; la modificación del mismo por Shi & Tomasi (conocida como *Good Features To Track detector*) [26] y la realizada por Triggs [27]; el detector FAST [28]; y CenSuRE (*STAR detector*) [29]. En los algoritmos modernos de extracción, se incluye tanto un detector como un descriptor, destacando las alternativas presentadas a continuación.

- **SIFT:** *Scale-invariant Feature Transform* [30]
  Permite extraer características invariantes a la iluminación, escala y orientación y robustas a cambios leves de perspectiva. Para lograr esto, primero se construye un espacio-escala piramidal construido mediante LoG (Laplacian of Gaussian), aproximado por un filtro de diferencias de gaussianas de distinto tamaño. Luego, se buscan esquinas que sean máximos locales y se asigna una orientación principal al punto de interés. Posteriormente, cada uno de estos puntos es descrito mediante vectores correspondientes a los histogramas de orientaciones de gradientes, obtenidos para una malla cuadriculada sobre la vecindad de los puntos.

- **SURF:** *Speeded Up Robust Features [31]*
  Es un extractor inspirado en SIFT, pero que en la etapa de detección aprovecha el uso de la imagen integral para calcular las diferencias de gaussianas en la búsqueda sobre el espacio-escala, sin necesidad de calcularlo completamente. En la descripción, se basa en la respuesta sobre filtros *Haar wavelets* para generar los vectores descriptores.

- **ORB:** *Oriented FAST and Rotated BRIEF* [32]
  Se presenta como una alternativa a SURF y SIFT, ofreciendo capacidades similares. Para detección implementa una adaptación al detector FAST para agregarle información sobre orientación de los puntos y al descriptor binario BRIEF, para añadir invariancia a la rotación. Al producir descriptores binarios se logra mayor velocidad en la extracción y comparación entre características.

- **KAZE [33]:**
  Presenta la novedad de utilizar un espacio-escala no lineal, construido a partir de técnicas de *Additive Operator Splitting* (AOS) para filtrado de difusión no lineal. Este enfoque permite que no se pierda resolución en los bordes de los objetos presentes en la imagen por aplicar el filtro gaussiano. Para cada punto de interés detectado, se calculan los descriptores *M-SURF* introducidos en [29] y adaptados al espacio-escala no lineal.

- **AKAZE:** *Accelerated-KAZE*
  El extractor de características AKAZE presenta el mismo enfoque de KAZE, pero acelera la construcción del espacio-escala no lineal usando técnicas *Fast Explicit Diffusion* (FED) [34] para las aproximaciones numéricas. Además, se cambia el descriptor por uno binario (M-LBD), aumentando la velocidad de cómputo [35].

- **ASIFT:** *Affine-SIFT* [36]
  Corresponde a una extensión del extractor SIFT, que añade mayor invariancia a transformaciones afines. Para lograr esto, primero se simulan las vistas posibles al variar la orientación del eje óptico de la cámara en los ángulos de latitud y longitud y luego, se computan los puntos SIFT para cada una de ellas, de manera de que quedan determinados todos los parámetros de la transformación afín en cada punto de interés. Al ampliar la dimensión de las características extraídas y requerir simulaciones de distintas vistas, el costo computacional es bastante mayor al de los otros extractores, pero presenta una gran mejora en sus resultados para asociar imágenes con alto grado de cambio de perspectiva.

- **DAISY:** *A Fast Local Descriptor for Dense Matching*
  Es un descriptor puro que se puede usar de manera densa sobre todos los puntos de la imagen. A grandes rasgos, el concepto es similar al de SIFT debido a que se basa en histogramas de orientaciones de gradientes. Sin embargo, al estar diseñado para estimar un mapa de profundidades, introduce mejoras en el cómputo de los histogramas, reduciendo el costo de los cálculos [37].

Como se mencionó anteriormente, también es posible obtener características sobre regio-

nes de una imagen, siendo el detector *Maximally Stable Extremal Region* MSER [38] una alternativa de este tipo. En [39], se realiza un estudio comparativo de múltiples detectores de regiones existentes, complementando lo expuesto en extracción de características.

2.2.3. Registro de imágenes

En el esquema planteado, una vez que se extraen una serie de características en cada imagen, es necesario establecer correspondencias entre las características que permitan asociar imágenes y determinar un conjunto consistente de transformaciones para alinearlas. Dependiendo de las características de la aplicación, puede ser necesario realizar un registro para cada par de imágenes y, a partir de esto, encontrar las transformaciones globales como ocurre en el caso de sistemas que deben operar a tiempo real [40]. De lo contrario, si se tiene disponibilidad de todas las imágenes al momento de realizar el registro, es posible hacerlo directamente de forma global, lo que usualmente se traduce en un costo computacional mayor.

Correspondencias de características

Dado un conjunto de características extraídas sobre puntos de interés de múltiples imágenes, esta etapa del algoritmo compara las características para encontrar las correspondencias existentes entre estos puntos, buscando reducir al mínimo la cantidad de calces incorrectos. Para ello, es importante utilizar una métrica de comparación acorde al descriptor utilizado. En el caso de descriptores binarios, se usa comúnmente la distancia de Hamming, mientras que en los otros casos la métrica más utilizada es la distancia Euclidiana. Si el conjunto de imágenes corresponde a una secuencia ordenada, es posible realizar encontrar las correspondencias entre pares de imágenes consecutivas, lo que, finalmente, permite relacionar toda la secuencia.

Existen variadas técnicas para establecer relaciones entre dos conjuntos de características en imágenes distintas, siendo las más relevantes detalladas en la Sección 4.1.3 de [5]. Según lo explicado por el autor, se puede utilizar directamente un algoritmo de calces por "fuerza bruta", comparando cada *feature* de una imagen con todos los existentes en la otra. Alternativamente, el algoritmo conocido como *K-Nearest Neighbors* (KNN), realiza un indexado tal que hace eficiente encontrar las K características más cercanas para la que se está comparando. En [41], se introduce una versión de KNN con capacidad de relacionar descriptores binarios. El hecho de tener más de un calce por característica hace posible eliminar calces incorrectos con un métodos heurístico que calcula el *Nearest Neighbor Distance Ratio* (NNDR), una suerte de medida de confusión establecida por la razón de distancias entre las dos características vecinas más cercanas, que permite filtrar correspondencias incorrectas fijando un umbral para esta medida.

Elección de transformación

La siguiente etapa del registro consiste en la búsqueda de las transformaciones geométricas que mejor representan las correspondencias conocidas, de manera de relacionar el contenido de las imágenes bajo un sistema de referencia global. Para ello, es importante realizar previamente una elección adecuada de las transformaciones que modelarán los cambios inducidos entre las imágenes, esto determinará si el registro es capaz de alinear todo el contenido de la imagen o sólo una parte de ella. En la literatura, se presentan distinciones para aplicaciones donde se trata de registrar una escena en la cual el contenido se encuentra principalmente en un plano o corresponde a una escena de mayor complejidad. Además, distinguen los casos donde la cámara realiza únicamente un movimiento de rotación o aquellos donde existe traslación que puedan producir efectos de paralaje sobre objetos cercanos.

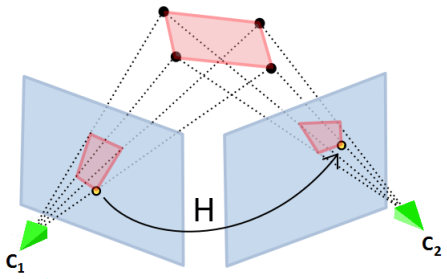


Figura 2.7: Transformación de homografía [42].

En primer lugar, en el caso de una escena plana o de una captura en que el movimiento de la cámara es de rotación pura, se pueden relacionar las distintas vistas de la escena mediante transformación de perspectiva en el espacio 2D, conocidas como **homografías**. Una homografía H establece una relación entre los puntos $\mathbf{x}$ y $\mathbf{x}'$ de dos vistas distintas en coordenadas homogéneas, como muestra la Ec. 2.1, definida para una escala indefinida en el espacio homogéneo. En la Fig. 2.7, es posible observar cómo la transformación H relaciona directamente, en el plano de las imágenes, los puntos correspondientes, sin necesidad de conocer el punto de la escena que es proyectado. Este es el enfoque más utilizado en los algoritmos tradicionales de stitching debido a que usualmente se aplican sobre imágenes de paisajes distantes que se aproximan a una escena plana [6].

$$\mathbf{x}' = \mathrm{H}\mathbf{x} \tag{2.1}$$

A modo de ejemplo, la Fig. 2.8 muestra el resultado obtenido al proyectar, mediante homografías, las dos imágenes laterales bajo movimiento de rotación, hacia el plano de la imagen central. Se observa que las zonas pertenecientes al plano central son proyectadas de manera coherente con el contenido, al contrario de aquellas zonas con componentes fuera del plano, las cuales presentan un alto grado de distorsión.

Investigaciones recientes buscan lograr mejores resultados en el registro de imágenes sobre escenas de mayor complejidad, aprovechando de forma más sofisticada las transformaciones homográficas. En [44], se propone el uso de homografías duales que permiten modelar la existencia de un plano lejano y un plano cercano y son utilizadas de manera ponderada. Una

Figura 2.8: Ejemplo de resultado obtenido al aplicar transformaciones de homografía globales para la alineación de imágenes [43].

aproximación más realista, basada en el trabajo anterior, es desarrollada en [45], donde se estiman múltiples homografías que varían suavemente dentro de una grilla. Si bien este último método logra obtener resultados más acertados sobre escenas complejas, para ajustarse bien a la escena requiere estimar un gran número de parámetros para la cantidad de homografías existentes, además de trabajar bajo el supuesto de que las regiones modeladas poseen planaridad local. La Fig. 2.9 permite contrastar los resultados que se pueden obtener con un método que incorpora homografías duales respecto de los que se obtendrían considerando sólo una única homografía para el plano distante o para el plano del suelo.

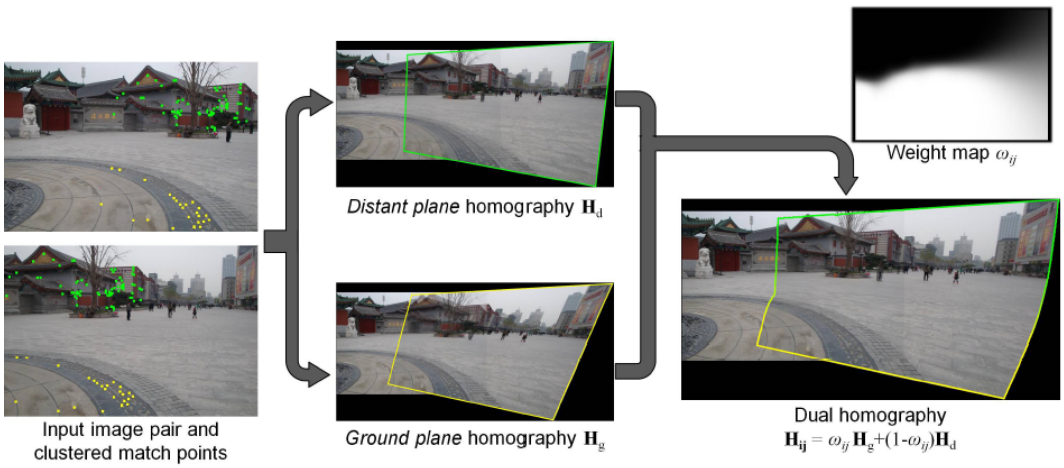


Figura 2.9: Ejemplo de resultado obtenido al aplicar transformaciones de homografías duales representando un plano lejano y uno correspondiente al suelo. [44].

En segundo lugar, cuando existen movimientos amplios de traslación y la escena no corresponde a un plano, los enfoques tradicionales buscan obtener el cambio de pose relativo entre cámaras, por lo que se debe encontrar una transformación rígida en el espacio 3D, también llamada **transformación euclidiana**, que represente este cambio. La Fig. 2.10 permite observar cómo se proyecta el punto $\mathbf{X}_j$ del mundo real al plano de la imagen correspondiente, proceso definido por las Ec. 2.2 y 2.3, donde se utiliza un sistema de referencia centrado en la cámara C_1 y una matriz de parámetros intrínsecos K conocida e igual para ambas cámaras.

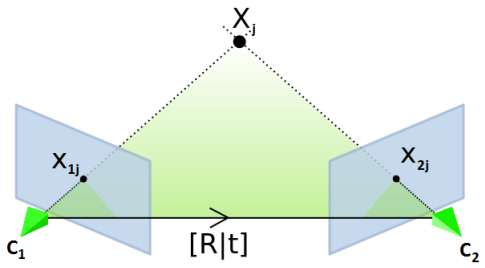


Figura 2.10: Transformación Euclidiana [46].

$$\mathbf{x}_{1j} = \mathrm{K}\,[\mathbb{I}_3|0]\,\mathbf{X}_j \tag{2.2}$$

$$\mathbf{x}_{2j} = \mathrm{K}\,[R|t]\,\mathbf{X}_j \tag{2.3}$$

En este contexto, la obtención de poses de las cámaras a partir de correspondencias en imágenes no es directa, ya que la relación que existe entre los puntos de imagen correspondientes se encuentra enmarcada en la geometría epipolar, que describe relaciones entre el espacio y dos cámaras proyectivas. Los detalles de la geometría epipolar son explicados en profundidad en la Parte II de [13]. La geometría epipolar, para el caso de dos cámaras, queda representada por la **matriz fundamental** F según la Ec. 2.4, aunque, si se conoce la matriz de parámetros intrínsecos de la cámara K, es posible utilizar directamente la **matriz esencial** E para tener una relación entre dos puntos de imagen correspondientes con el punto de la escena según la Ec. 2.6, mediante lo que se conoce como la restricción de coplanaridad, que representa la pertenencia de los puntos de la imagen al plano epipolar, delimitado por los centros de proyección de cada cámara y el punto de la escena proyectado.

$$(\mathbf{x}')^T \mathrm{F}\mathbf{x} = 0 \tag{2.4}$$

$$\mathrm{E} = \mathrm{K}^T \mathrm{F}\mathrm{K} \tag{2.5}$$

$$(\mathbf{x}')^T \mathrm{E}\mathbf{x} = 0 \tag{2.6}$$

En la Fig. 2.11, se muestra el resultado obtenido para el caso presentado en la Fig. 2.8, pero considerando transformaciones euclidianas de rotación para relacionar las imágenes. Se puede observar que la imagen resultante es más consistente con la escena real y presenta menores distorsiones de perspectiva respecto a las obtenidas al usar homografías globales.

Basándose en los distintos tipos de movimiento explicados en la Sección 2.2.1, existen trabajos que imponen ciertas restricciones sobre el movimiento para poder facilitar la estimación de las transformaciones euclidianas. En [47], se trata el caso donde se asume rotación pura de la cámara, mientras que en [48], se restringe una dirección vertical conocida asumiendo que naturalmente las fotografías son tomadas de esta forma. Estas simplificaciones reducen el espacio de búsqueda para los parámetros de las transformaciones y disminuyen costos de cómputo, además de asegurar mayor estabilidad en las estimaciones. Sin embargo, se tiene la desventaja de que el usuario debe asegurarse de cumplir las restricciones de movimiento al realizar la captura.

Figura 2.11: Ejemplo de resultado obtenido al aplicar transformaciones euclidianas de rotación para la alineación de imágenes [43].

Estimación de homografías

La estimación de la transformación suele ser abordada de manera robusta mediante el algoritmo iterativo *RANdom SAmple Consensus* (RANSAC), que permite encontrar un subconjunto de correspondencias validadas contra este modelo [49]. El método *Direct Linear Transform* (DLT) es usualmente utilizado para resolver el sistema de 2.1 y encontrar una matriz H en cada iteración de RANSAC. Además, es posible utilizar este método sobre un conjunto de más de 4 correspondencias, cuando el sistema está sobre-determinado, obteniendo una solución de tipo *least-squares* [50].

Estimación de poses de cámara

La estimación de las transformaciones euclidianas que definen las poses de cámaras a partir de la información visual es un problema general que trasciende su aplicación en algoritmos de stitching, siendo estudiado profundamente en tópicos de visión por computadora como calibración estéreo [51], reconstrucción 3D de escenas por *Structure from Motion* [52] o *Multi-View Stereo* [53], navegación de vehículos autónomos por odometría visual [16] y realidad aumentada [54], entre otros.

Dado que la relación entre los puntos correspondientes en un par de imágenes está comprendida en la matriz esencial, la estimación de pose requiere estimar, primero, esta matriz, lo cual necesita un conocimiento previo de los parámetros intrínsecos de la cámara [55]. Luego, mediante descomposición de esta matriz, se puede estimar la rotación y la dirección de la traslación. La magnitud de la traslación mantiene un factor de escala desconocido debido a la geometría proyectiva del problema, por lo que, para completar esta estimación, se requiere incorporar información adicional sobre la estructura de la escena, como por ejemplo, el tamaño real de un objeto proyectado en la imagen [56],[57]. Un problema relacionado en esta área es conocido como *Perspective-n-Point* (PnP) que consiste en estimar la pose de una cámara a partir del conocimiento de correspondencias entre puntos 3D de la escena y sus proyecciones en la imagen [58].

Reconstrucción 3D de escenas

Al conocer las poses de la cámara y las correspondencias entre proyecciones en la imagen de algunos puntos de la escena, es posible estimar la posición en el espacio de estos puntos. Si bien para los algoritmos de stitching tradicionales esto no es un tópico de interés, existen trabajos donde se realiza la fusión de imágenes sobre modelos tridimensionales de la escena [7], por lo que ampliando el concepto a este tipo de trabajos, la reconstrucción de escenas puede incorporarse como una etapa más del stitching que utiliza registro de poses de cámara.

El problema de reconstrucción de escenas posee variados métodos de resolución, principalmente diferenciados por la cantidad de cámaras involucradas y la disponibilidad de las múltiples vistas al momento de realizar la reconstrucción. Por un lado, aquellos algoritmos que utilizan una cámara en movimiento y que pueden ser implementados en tiempo real, pertenecen al área conocida como *Structure From Motion* (SFM) [52]. Por otro lado, aquellos que poseen la información simultánea de dos o más vistas, desde una o más cámaras, pertenecen a las ramas de investigación conocidas como *Stereo Vision*[51] o *Multi-view Stereo* (MVS) [53].

Estos algoritmos, utilizan la triangulación de los puntos de imagen correspondientes en distintas vistas para obtener una posición del punto de la escena que representan. El problema de triangulación está definido en [59], donde se recopilan y comparan los distintos algoritmos para resolverlo, siendo el más directo el que utiliza la *Direct Linear Transform* (DLT), que se extiende fácilmente al uso de múltiples vistas.

Una vez que se posee una estimación de las poses de cámara y puntos de la escena, con sus proyecciones respectivas en cada vista, los algoritmos usualmente realizan una etapa de optimización que refina esta reconstrucción inicial. En el caso de MVS, tanto la estimación como la optimización se realizan de manera simultánea, permitiendo una mayor precisión en la reconstrucción final, pero con altos costos de cómputo. Por otro lado, el SFM posee un flujo secuencial que comienza con una estructura inicial creada a partir de triangulación desde las dos primeras vistas y, posteriormente, agrega nuevas vistas y puntos a la reconstrucción, permitiendo, cada cierto número de vistas agregadas, optimizar la reconstrucción. Este enfoque es útil para aplicaciones que requieren funcionar en tiempo real, pero significan una pérdida en la precisión final. En el Capítulo 7 de [5], se describe a profundidad el tópico de los algoritmos de SFM, mientras que en [13] se detallan las técnicas de MVS, incluyendo también los casos particulares de reconstrucción desde una, dos y tres vistas de una escena.

En el ámbito de la evaluación de algoritmos de reconstrucción de escenas, se destacan los trabajos [60] y [61], que presentan bases de datos diseñadas para este propósito. En [61], el autor construyó múltiples bases de datos para la evaluación de algoritmos de MVS, presentando secuencias de imágenes de alta resolución, que capturan distintas escenas de edificaciones complejas desde puntos de vista con amplios desplazamientos de cámara. En estas bases de datos, las imágenes han sido previamente rectificadas, corrigiendo los efectos de distorsión radial del lente. Adicionalmente, en ambos trabajos, se provee la matriz de parámetros intrínsecos de la cámara utilizada para cada secuencia.

Bundle Adjustment

En el mundo de la fotogrametría y reconstrucción de escenas mediante información visual, la técnica conocida como *Bundle Adjustment* (BA), es una etapa final de optimización compartida por casi todos los algoritmos existentes en la materia. En esta etapa, se busca refinar los parámetros de las transformaciones encontradas, minimizando el error cuadrático total de reproyección de los puntos de la escena a las múltiples vistas [62], [63]. Una versión análoga puede aplicarse para el caso en que la cámara posee únicamente movimientos de rotación pura, realizando la optimización en base a minimizar los ángulos entre haces de luz que describen el mismo punto para diferentes imágenes. Para este problema de optimización, se suele utilizar el algoritmo de Levenberg-Marquardt, como se muestra en [47]. Otros algoritmos de optimización usados en esta etapa se encuentran descritos en el Capítulo 7 de [5].

2.2.4. Composición de imágenes

Teniendo definidos los modelos que relacionan las imágenes, ya sea a través de homografías o transformaciones de las cámaras, ya es posible llevar las imágenes en un sistema de referencia común. Así, la etapa de composición consiste, en primera instancia, en proyectar las distintas vistas y, luego, escoger qué partes de cada imagen son incluidas en el contenido final. Este último proceso es conocido como *seam cutting* o elección de bordes y afecta considerablemente la calidad de la composición final obtenida. Además, para que los bordes no sean visibles, hay aplicaciones que incluyen una etapa llamada *blending*, encargada de mezclar el contenido traslapado de las imágenes que se están componiendo. En la Sección 9.3 de [5], se describen los métodos más utilizados para la composición de imágenes, al igual que el marco teórico correspondiente.

Para la etapa de proyección existen diversas alternativas, las cuales difieren principalmente en la superficie elegida para proyectar las imágenes. Las opciones más comunes incluyen superficies como un plano, un cilindro o una esfera [64]. Sin embargo, en [65] se propone la proyección sobre una superficie parametrizada de una manera tal, que reduce las distorsiones de perspectiva en la imagen final, haciendo la construcción de la superficie parte del proceso de optimización realizado al registrar las imágenes.

En este mismo enfoque, si el registro involucra obtener una reconstrucción parcial de la escena es posible proyectar el contenido de las imágenes sobre una representación aproximada de la escena ajustando superficies B-Spline, como se realiza en [66]. En este caso, al contar con una malla de polígonos, se puede usar un método de proyección como el de [7], donde se mezclan imágenes sobre un modelo 3D, considerando la proyección sobre las caras de la superficie que enfrentan a la cámara desde donde se proyecta. Es en este mismo trabajo, al igual que en [67], donde se demuestra que una elección inteligente de bordes puede facilitar e incluso evitar por completo la etapa de *blending*. El enfoque utilizado por el autor utiliza los algoritmos *watershed segmentation* [19] y optimización de grafos *graph-cut*, similar a la utilizada en [68], para construir de manera coherente el mosaico final, a partir de segmentos de cada imagen individual.

Finalmente, al tener alineadas las imágenes proyectadas sobre la misma superficie y con sus zonas de traslape definidas, la información en estas zonas puede combinarse desde las distintas fuentes para evitar dejar un borde notorio en casos donde existen diferencias considerables en exposición. Como es descrito en la Sección 9.3 de [5], las técnicas de *blending* existentes van desde el *feathering*, que realiza interpolación de los valores de intensidad, hasta *multi-band blending*, que mezcla la información en distintas bandas de frecuencia en base a una pirámide gaussiana.

2.3. Herramientas

Además de la teoría existente en el campo de visión por computadoras y algoritmos de stitching, es necesario revisar algunas de las herramientas comúnmente utilizadas para desarrollar los sistemas que hacen uso de las técnicas existentes en el área. En este contexto, se identifica la existencia de la librería *Open Computer Vision* (OpenCV) [69], creada para resolver problemas de visión computacional, la cual incluye módulos específicos para el procesamiento de imágenes, extracción de características y estimación de homografías y cambios de pose, entre otros. Otra librería a destacar es *Point Cloud Library* (PCL) [70], que permite manipular fácilmente nubes de puntos 3D e implementa un módulo de ajuste de superficies [71]. Finalmente, se destaca también la librería de *Sparse Bundle Adjustment* (SBA) [72], que implementa eficientemente el algoritmo de *Bundle Adjustment* mediante optimización por el método de Levenberg-Marquardt. En el ámbito de herramientas de visualización de reconstrucciones tridimensionales, destaca el software llamado *Meshlab* [73], que permite visualizar tanto nubes de puntos, con y sin color, como superficies, entre otras funcionalidades.

Capítulo 3

Metodología

En este capítulo, se presenta el diseño del algoritmo propuesto para la resolución del problema planteado, profundizando en las técnicas utilizadas para cada etapa. Además, se plantea el plan de trabajo para desarrollar el diseño propuesto y se definen las métricas utilizadas para la evaluación.

3.1. Método

De acuerdo al estudio realizado sobre los algoritmos de stitching, se plantea utilizar como base el diseño general presentado en el capítulo anterior, en la Fig. 2.2, que aborda el problema de fusión de imágenes para el caso de escenas estáticas a partir del registro mediante extracción y calce de puntos de interés. Para aplicar este diseño a las secuencias del *Logmeter*, se realiza una adaptación sencilla, que será explicada en esta sección.

Así, el diseño propuesto está compuesto por una serie de bloques independientes de procesamiento, que siguen el flujo mostrado en el diagrama de la Fig. 3.1 y otorgan flexibilidad en la implementación de cada uno. En particular, el bloque de alineamiento y fusión de imágenes puede ser implementado tanto por registro de homografías, como por registro de poses de cámara, donde el último involucra una reconstrucción parcial de la escena. A continuación, se describen, en primera instancia, los algoritmos utilizados para las primeras tres fases, para posteriormente detallar el enfoque abordado para el desarrollo de la parte distintiva de los algoritmos de stitching: el alineamiento y fusión de imágenes.

3.1.1. Pre-procesamiento

Para resaltar los detalles en las imágenes, se propone utilizar la mejora de contraste, mediante la técnica CLAHE [18], que realiza ecualización de histogramas locales sobre una imagen en escala de grises. Esto permitiría una mejor extracción de características, aumentando la cantidad y calidad de los descriptores que permitirán establecer correspondencias.

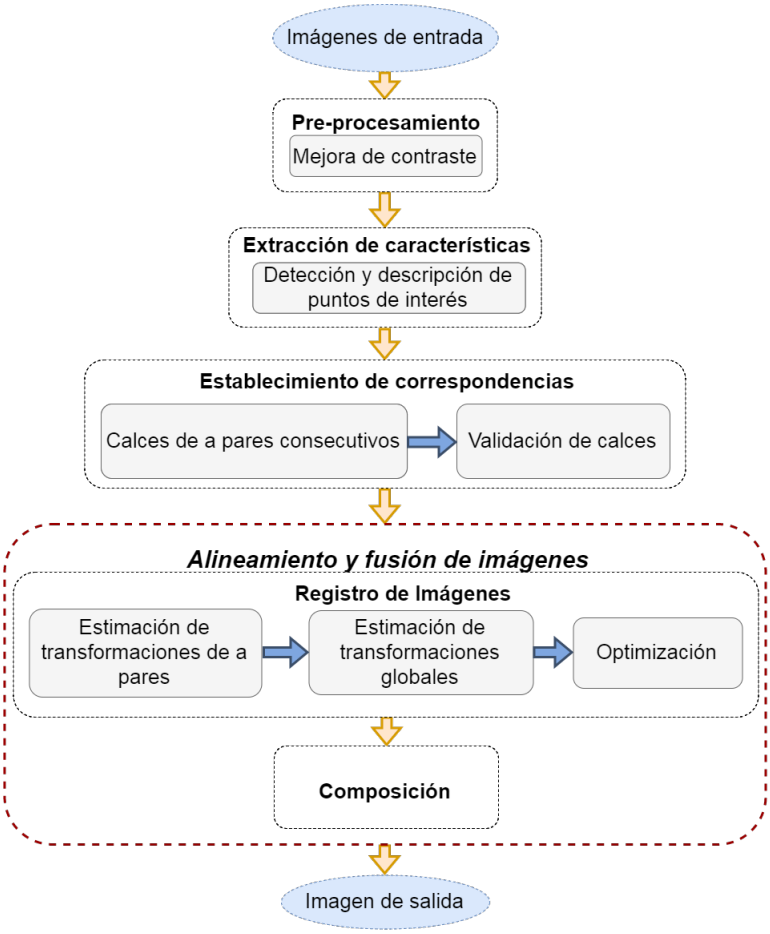


Figura 3.1: Diagrama global del algoritmo de stitching propuesto.

Mejora de contraste CLAHE

Este método consiste en dividir la imagen en regiones un tamaño definido, en las cuales se realiza una ecualización de histograma, fijando una cota superior a la amplificación de contraste en el interior de la región. Esta ecualización de histograma, realiza una transformación sobre la intensidad de los pixeles, de manera de obtener una función de distribución acumulativa (c.d.f.) cercana a una recta, como se muestra en la Fig. 3.2. La función que define la transformación sobre una región de tamaño $M \times N$ es descrita en la Ec. 3.1, donde x es el valor de intensidad del pixel, $\text{cdf}(x)$ es la función de distribución acumulativa, asumiendo una discretización de la intensidad en un rango de $[0, 255]$.

$$h(x) = \left\lfloor \frac{\text{cdf}(x) - \text{cdf}_{min}}{(M \cdot N) - 1} \cdot 255 \right\rfloor \tag{3.1}$$

La amplificación de contraste está directamente relacionada con la pendiente de la recta obtenida para la c.d.f. luego de realizar la ecualización. En CLAHE, se limita esta amplificación para no aumentar el ruido, mediante una redistribución de los pixeles que superen un umbral en el histograma original de la imagen, como se muestra en la Fig. 3.3, calculando sobre el histograma resultante la función de distribución acumulativa utilizada para la ecualización.

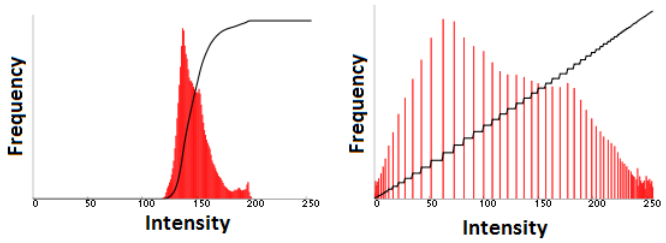


Figura 3.2: Histograma y curva de función de distribución acumulativa antes y después de la ecualización del histograma.

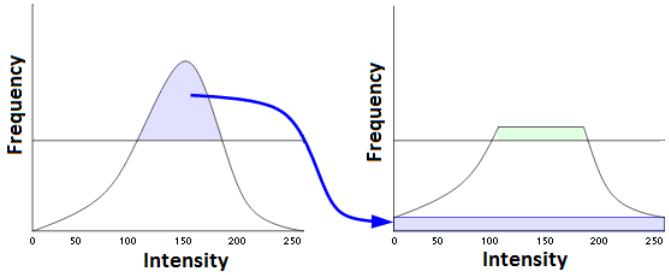


Figura 3.3: Limite de corte para acotar amplificación de contraste.

3.1.2. Extracción de características

El extractor de características escogido para esta etapa del algoritmo es AKAZE [35], que, como muestra el estudio realizado por el autor, destaca entre sus pares manteniendo un desempeño similar a SIFT [30], pero con menores costos computacionales de almacenamiento y comparación, ya que utiliza un descriptor binario.

AKAZE

El algoritmo de extracción utilizado en AKAZE permite obtener características invariantes a intensidad, orientación y escala. Para lograr esto, al igual que en SIFT [30], se comienza construyendo una representación multi-escala de la imagen, lo cual permite la detección y descripción de puntos de interés a distintas escalas. El principio básico tras este procesamiento consiste en crear un espacio donde se varía la escala filtrando la imagen original con una función apropiada para cada nivel.

En AKAZE, como se describe en [33], se utiliza un filtro de difusión no lineal, basado en modelar la evolución de la intensidad de una imagen a través de niveles de escala como la divergencia de una función de flujo que controla el proceso de difusión al aumentar la escala. Esto permite filtrar manteniendo los bordes de los objetos y, por ende, la localización precisa de puntos de interés, al contrario de lo que ocurre utilizando filtros gaussianos de distinto tamaño, como se hace en SIFT [30]. La Fig. 3.4, a modo de ejemplo, compara el efecto del filtro de difusión no lineal con el filtrado gaussiano equivalente en tres escalas distintas, permitiendo observar el efecto mencionado.

Figura 3.4: Comparación de filtro de difusión no lineal, en la parte superior, y filtro gaussiano, en la inferior, para niveles de escala equivalentes.

El flujo de difusión no lineal de intensidad lumínica, para una imagen L, está modelado por la ecuación en derivadas parciales Ec. 3.2, donde div y ∇ corresponden a los operadores de divergencia y gradiente, respectivamente. La variable t representa la escala y c es una función de conductividad que permite controlar la variación de la difusión de acuerdo a la estructura local en la imagen y la escala. En este caso, se considera una conductividad de difusión variable en función de la magnitud de la imagen gradiente para cada nivel de escala. La creación del espacio escala no lineal en la implementación de AKAZE, es realizada mediante el esquema *Fast Explicit Diffusion* (FED) [34], que resuelve numéricamente el sistema de la Ec. 3.2 para aplicar el filtrado de difusión no lineal.

$$\frac{\partial L}{\partial t} = \text{div}(c(x, y, t) \cdot \nabla L) \tag{3.2}$$

Por otro lado, la detección de los puntos de interés es realizada calculando, en todos los niveles de escala, la función de respuesta con el determinante del Hessiano, la cual es normalizada según el nivel de escala. Posteriormente, se realiza una búsqueda de máximos locales en el espacio conjunto de imagen y espacio escala, que definen la posición de los puntos de interés, con una orientación asociada.

Finalmente, conociendo la ubicación de los puntos de interés, la descripción de su vecindad se realiza con el descriptor *Local Difference Binary* (M-LDB). Este método, divide cada vecindad en una grilla, alineada con la orientación del punto de interés. Dentro de cada celda de la grilla, se calcula el promedio de la intensidad y el gradiente en la dirección horizontal y vertical. Después, se realizan comparaciones binarias de los valores calculados entre pares de celdas. Para que el descriptor sea invariante ante cambios de escala, se realizan múltiples divisiones de la vecindad, cambiando el tamaño de la grilla, además de realizar un submuestreo de los pixeles contenidos en cada celda, de tamaño proporcional a la escala, con el objetivo de reducir los costos de cómputo del descriptor. Los resultados de las comparaciones entre

celdas son concatenados en un vector descriptor binario, el cual caracteriza el punto de interés y permite identificarlo en otras imágenes, incluso bajo cambios de escala, rotación e iluminación.

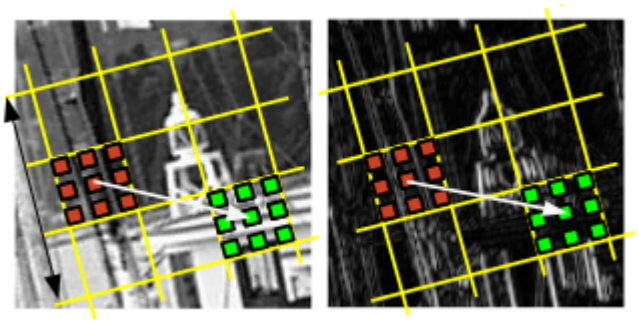


Figura 3.5: Cálculo de descriptor *Modified-Local Difference Binary* (M-LDB), para un nivel de escala.

3.1.3. Establecimiento de correspondencias

Conociendo el conjunto de puntos de interés con sus descriptores extraídos en cada vista, se decide buscar correspondencias entre todos los pares de imágenes consecutivas, bajo el esquema de *K-Nearest Neighbors* para descriptores binarios [41]. Primero, se hace un filtrado inicial mediante un umbral de distancia correspondiente a tres veces la distancia mínima entre los pares de descriptores, permitiendo descartar aquellos que, relativamente, son lejanos. Luego, para cada descriptor, se encuentran los dos descriptores vecinos más cercanos, permitiendo realizar una etapa de filtrado mediante imposición de un umbral sobre la medida NNDR descrita en [30], donde se propone usar valor de 0,8.

El filtrado por NNDR se basa en que cada calce correcto debe estar a una distancia significativamente menor que la del calce incorrecto más cercano. Así, el indice NNDR es calculado como la razón entre las distancias de un descriptor $\vec{\mathrm{d}}$ a los dos descriptores vecinos más cercanos $\vec{\mathrm{d}}_{1^{st}NN}$ y $\vec{\mathrm{d}}_{2^{nd}NN}$, como se muestra en la Ec. 3.3, donde se impone la restricción del umbral para eliminar correspondencias incorrectas.

$$\mathrm{NNDR}(\vec{\mathrm{d}}) = \frac{\mathrm{dist}(\vec{\mathrm{d}}, \vec{\mathrm{d}}_{1^{st}NN})}{\mathrm{dist}(\vec{\mathrm{d}}, \vec{\mathrm{d}}_{2^{nd}NN})} < 0{,}8 \tag{3.3}$$

Estimación de homografías

Una homografía define una transformación de perspectiva en el espacio 2D, por lo que puede representar el cambio existente entre dos imágenes planas. Esta transformación, es expresada en una matriz cuadrada de 3×3, la cual se normaliza fijando el último elemento al valor 1, ya que, en el espacio de coordenadas homogéneas, la misma transformación de perspectiva puede ser representada en múltiples factores de escala. Así, la homografía queda

completamente definida por 8 parámetros y se requiere un mínimo de 4 puntos correspondientes para poder estimarlos.

$$\mathrm{H} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix} \tag{3.4}$$

Para realizar una estimación robusta de la transformación, se utiliza RANSAC [49], que, de manera iterativa, selecciona un subconjunto aleatorio de cuatro correspondencias para generar una hipótesis de homografía H y, posteriormente, determina la cantidad de correspondencias que son consistentes con esta hipótesis, valor conocido como el **consenso**. Para decidir si un calce de puntos $\mathbf{x}_\mathrm{i}$ y $\mathbf{x'}_\mathrm{i}$ está en consenso con la hipótesis H, se utiliza el umbral definido en la Ec. 3.5, que define la distancia máxima aceptable entre la reproyección de un punto mediante la homografía y el punto correspondiente en la otra imagen. Finalmente, al realizar múltiples iteraciones, se acepta como válida la hipótesis que genere el mayor consenso, lo que da robustez a la estimación en base a la probabilidad de generar una buena hipótesis, ya que, de existir algunas correspondencias incorrectas, y generar una hipótesis de homografía con ellas, se tendrá un bajo consenso. Por otro lado, las hipótesis provenientes de los calces correctos son más probables y, además, contarán con un mayor consenso.

$$||\mathbf{x'}_\mathrm{i} - \mathrm{H} \cdot \mathbf{x}_\mathrm{i}|| \leq \mu_{RANSAC} \tag{3.5}$$

La generación de cada hipótesis es obtenida como una solución *least-square* que minimiza el error de reproyección cuadrático total, para las correspondencias aceptadas, como se muestra en la Ec. 3.6.

$$\min_{\mathrm{H}} \sum_{\mathrm{i}=1}^{n} ||\mathbf{x'}_\mathrm{i} - \mathrm{H} \cdot \mathbf{x}_\mathrm{i}||^2 \tag{3.6}$$

Validación de calces

Bajo el supuesto de que los objetos representados en la escena no sufren grandes deformaciones, se utiliza como etapa de validación de calces la estimación de una homografía con un umbral μ_{RANSAC} lo suficientemente grande para descartar los calces incorrectos que hayan superado la etapa de filtrado inicial por NNDR, imponiendo que las correspondencias estén en consenso con una geometría global común, independiente de si la escena es plana o no.

Adicionalmente, se considera una etapa de validación final, como la usada en [40], donde se validan las correspondencias establecidas al estimar de manera robusta las transformaciones en la etapa de alineamiento, aceptando aquellas correspondencias que estén de acuerdo con la transformación. La metodología propuesta considera esta última fase de validación por RANSAC, tanto en el registro por homografías, como en el caso del registro de poses de cámara. En el caso de estimación por homografías, la validación depende del umbral definido

para aceptar una hipótesis de homografía, por lo tanto, los calces validados corresponden a aquellos que están en consenso con la homografía aceptada en la estimación. Por otro lado, la validación de calces en el caso de registrar poses de cámara, ocurre al estimar la matriz esencial mediante RANSAC, etapa que será detallada más adelante y que, básicamente, utiliza los mismos principios que el caso anterior.

3.1.4. Metodología para el alineamiento y fusión de imágenes

Según lo estudiado, se identificó que los algoritmos de stitching presentan principalmente dos alternativas de transformaciones para el registro de imágenes: homografía o el de poses de cámara. Si bien el registro homográfico es comúnmente utilizado sobre escenas estáticas de un plano distante o con un movimiento de cámara puramente rotacional, se investigaron trabajos como [45] y [44] que logran aplicar homografías locales sobre escenas complejas con traslación de cámara.

Por este motivo, la metodología abordada considera una etapa inicial de exploración mediante el desarrollo de un prototipo que realice stitching por homografías globales, permitiendo probar las limitaciones reales de este método sobre las imágenes del *Logmeter*, además de conceder una evaluación temprana del desempeño de las etapas anteriores al registro de imágenes. Luego de evaluar los resultados de la fase de exploración, se presentan dos opciones para continuar el desarrollo:

1. Desarrollo del algoritmo de stitching homográfico, utilizando el registro local de [45].
2. Implementación del alineamiento mediante registro por transformaciones euclidianas, logrando un modelo 3D sobre el cual se puedan proyectar imágenes como en [66] o [7].

En este punto decisivo se toma la decisión de implementar la segunda opción, motivado por el valor adicional de una reconstrucción tridimensional, que permitiría realizar la inspección visual de la escena desde una vista arbitraria, además de otorgar, en la aplicación del producto de Woodtech S.A., la posibilidad de complementar el escaneo láser con información de las imágenes. A continuación, se detalla la metodología abordada para la etapa de alineamiento y fusión de imágenes, primero para el prototipo que hace registro mediante homografías y luego para el algoritmo de stitching por reconstrucción de escena y fusión de imágenes sobre una superficie 3D.

3.1.5. Algoritmo de stitching por homografías globales

En este enfoque, se alinean todas las imágenes mediante homografías globales hacia una imagen que define un plano de referencia. Las homografías globales son construidas mediante concatenación de las homografías obtenidas entre cada par consecutivo de imágenes, estimadas con RANSAC [49]. La composición de las homografías relativas se realiza multiplicando las matrices respectivas, según la Ec. 3.7, donde $H_{\mathrm{i},j}$ es la homografía que lleva el contenido de la imagen i a la imagen j, $\mathbb{I}_3$ es la matriz de identidad 3×3 y $H_{\mathrm{i}} = H_{\mathrm{i,ref}}$ es la homografía global definida para alinear la imagen i en el plano de referencia de la imagen ref.

$$H_{\mathrm{i}} = \begin{cases} H_{\mathrm{i,i+1}} \dots H_{\mathrm{ref-1,ref}} & \mathrm{i} < \mathrm{ref} \\ \mathbb{I}_3 & \mathrm{i} = \mathrm{ref} \\ H_{\mathrm{i,i-1}} \dots H_{\mathrm{ref+1,ref}} & \mathrm{i} > \mathrm{ref} \end{cases} \tag{3.7}$$

Cuando ya se conoce la alineación entre las imágenes, debido a que el enfoque de este trabajo está en la etapa de registro, se definen dos alternativas de fusión, una composición sencilla por superposición y una mediante *blending* lineal. Para la primera, inicialmente, se toma la última imagen de la secuencia y se proyecta directamente la zona no traslapada en la composición final. Luego, para cada par consecutivo de imágenes se determina la zona de traslape según la Ec. 3.8, proyectando únicamente el contenido de la imagen que está más hacia el final de la secuencia, si es que los pixeles no se encuentran ya coloreados. Finalmente, se proyecta la zona no traslapada de la primera imagen, completando la composición. En el caso de la segunda alternativa, el esquema es el mismo, salvo que se mezclan, para cada posición, todos los pixeles proyectados en las zona de traslape, mediante la Ec. 3.9, donde se consideran m imágenes proyectadas sobre el pixel x, $\mathrm{C}(x)$ es el valor final del pixel en la composición y $\mathrm{I_i}(x)$ es el valor de intensidad del pixel en la imagen i proyectada. Este método es descrito con mayor detalle en la sección 9.3.2 de [5].

$$\mathrm{Overlap}_{ij} = \mathrm{Mask_i} \cap \mathrm{Mask}_j \tag{3.8}$$

$$\mathrm{C}(x) = \frac{\sum_{\mathrm{i}}^{m} \mathrm{I_i}(x)}{m} \tag{3.9}$$

Como se verá en la Sección 4.3.3, el stitching por homografías posee fuertes limitaciones para ser aplicado sobre secuencias con amplios movimientos de cámara, por lo que se aborda, además, una metodología alternativa para el bloque de alineamiento y fusión de imágenes de la Fig. 3.1, la cual es descrita a continuación.

3.1.6. Algoritmo de stitching por reconstrucción de escena

En esta propuesta, se incluye una etapa que registra las poses de las cámaras mediante transformaciones euclidianas y logra una reconstrucción parcial de la escena, la cual es extendida a un modelo aproximado mediante ajuste de una superficie 3D. Finalmente, la fusión de imágenes se realiza sobre esta superficie en el espacio tridimensional.

Modelo de cámara *pinhole*

La propuesta considera un modelo de cámara *pinhole* para modelar el mecanismo de proyección involucrado en cada una de las vistas. En esta implementación, para el caso de las secuencias de *Logmeter*, se evalúa el desempeño de la matriz de calibración intrínseca con los parámetros extraídos desde los datos conocidos del sensor de la cámara (Ver Anexo 6.1). Además, se deja abierta la posibilidad para que en el etapas posteriores de desarrollo,

se incluya un paso previo de calibración que permita incluso generar un modelo de distorsión del lente para mejorar los resultados obtenidos compensando este efecto.

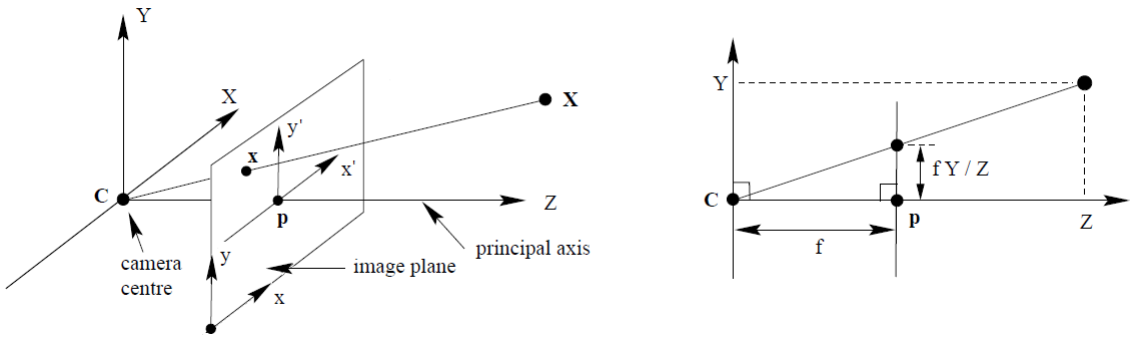


Figura 3.6: Geometría de modelo de cámara *pinhole* [13].

En la Fig. 3.6, se muestra la proyección de un punto $\mathbf{X}$ del mundo real al punto $\mathbf{x}$ del plano de la imagen ubicado en $z = f$, para el caso en que el centro proyectivo de la cámara $\mathbf{C}$ se encuentra en el origen de las coordenadas del mundo real. En un sistema global, el centro proyectivo se encuentra ubicado en $\tilde{C}$, por lo que basta aplicar una transformación euclidiana $[R|t]$ para llevar el punto al sistema coordenado con origen en la cámara, con R una matriz de rotación y $t = -R\tilde{C}$ un vector traslación. Así, la Ec. 3.10 representa la relación establecida entre la escena y una vista de la cámara *pinhole*, donde $\mathbf{X}$ es el punto del mundo real y $\mathbf{x} = (x, y, 1)^T$ es el punto proyectado al sistema de referencia de la imagen, en coordenadas homogéneas. La transformación proyectiva $\mathrm{P} = \mathrm{K}R[\mathbb{I}_3|-\tilde{C}]$ es conocida como la matriz de proyección de la cámara.

$$\mathbf{x} = \mathrm{K}R[\mathbb{I}_3|-\tilde{C}]\mathbf{X}$$

$$K = \begin{bmatrix} f & 0 & x_0 \\ 0 & \alpha f & y_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.10}$$

En esta ecuación, K es la matriz de parámetros intrínsecos de la cámara, los cuales incluyen la **longitud focal** horizontal f, la traslación hacia el punto principal $\mathbf{p} = (x_0, y_0)$ y un factor α correspondiente a la razón de aspecto para la longitud focal vertical. Mientras que la transformación euclidiana $[R|t]$ conforma los parámetros extrínsecos de la cámara.

Estimación de matriz esencial y poses de cámara

La matriz esencial que impone la restricción de coplanaridad de la Ec. 2.6 sobre los puntos de la escena y sus proyecciones en las distintas imágenes, se define según la Ec. 3.11, donde $[t]_x$ representa la forma matricial del producto cruz. En este sentido, se puede estimar la rotación y la dirección de la traslación, pero la magnitud de esta mantiene un factor de escala desconocido.

$$
\begin{aligned}
\mathrm{E} &= [t]_x R \\
[t]_x &= \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}
\end{aligned} \tag{3.11}
$$

Conociendo las correspondencias establecidas entre pares consecutivos de imágenes, al igual que en el caso de registro homográfico, se utiliza el método de estimación robusta RANSAC para eliminación de calces incorrectos y obtener la matriz esencial E que mejor representa la geometría epipolar del conjunto de datos. En cada iteración se utiliza el método [55] que requiere un mínimo de 5 correspondencias para estimar la matriz esencial.

Para la extracción de la pose relativa de las cámaras, se extrae la matriz de rotación R y la dirección de la traslación $\vec{t}$, de la matriz esencial mediante el algoritmo descrito en [55]. Se comienza por una descomposición *Singular Values Decomposition* (SVD) de la matriz, la cual permite obtener 4 posibilidades para la geometría del par de cámaras. Para escoger la correcta, se realiza la triangulación de los puntos desde las cámaras obtenidas para verificar que la geometría escogida sea aquella en que los puntos de la escena se encuentren frente a ambas cámaras, esta etapa de verificación es conocida como *cheirality check*. Es importante recalcar que en esta estimación se desconoce la magnitud de las traslaciones, por lo que el registro inicial no es exacto y requiere la etapa posterior de optimización.

Posterior a la estimación de poses relativas, se combinan las transformaciones euclidianas de cada par consecutivo para obtener una transformación global para cada cámara utilizando como referencia la pose de la cámara de la primera vista. Para la combinación de poses, se sigue el esquema recursivo de la Ec. 3.12, donde $\mathrm{T}_{\mathrm{i},j}$ es la transformación que representa la pose relativa de la cámara j respecto a la pose de la cámara i, T_i corresponde a la pose global de la cámara i.

$$
\begin{aligned}
\mathrm{T}_\mathrm{i} &= \mathrm{T}_{\mathrm{i}-1} \cdot \mathrm{T}_{\mathrm{i}-1,\mathrm{i}} \\
\mathrm{T}_0 &= [\mathbb{I}_3 | \mathbf{0}]
\end{aligned} \tag{3.12}
$$

3.1.7. Reconstrucción 3D de escena

El diseño propuesto para este algoritmo de stitching, se basa en resolver el problema de reconstrucción de escena presentado en la Sección 2.2.3. Para ello, se realiza una reconstrucción parcial de escena inicial, construida a partir de las estimaciones de cámaras y la triangulación de las correspondencias conocidas, utilizando el método por *Direct Linear Transform* (DLT) [59]. Esta reconstrucción inicial, es optimizada mediante *Sparse Bundle Adjustment* global [72], aprovechando la disponibilidad simultánea de todas las imágenes de la secuencia. Finalmente, como ilustra la Fig. 3.7, se obtiene una superficie sobre la cual proyectar las vistas de cada cámara, mediante el ajuste de una superficie B-Spline cúbica a la nube de puntos [66]. El flujo de la reconstrucción de escena es resumido en la Fig. 3.8,

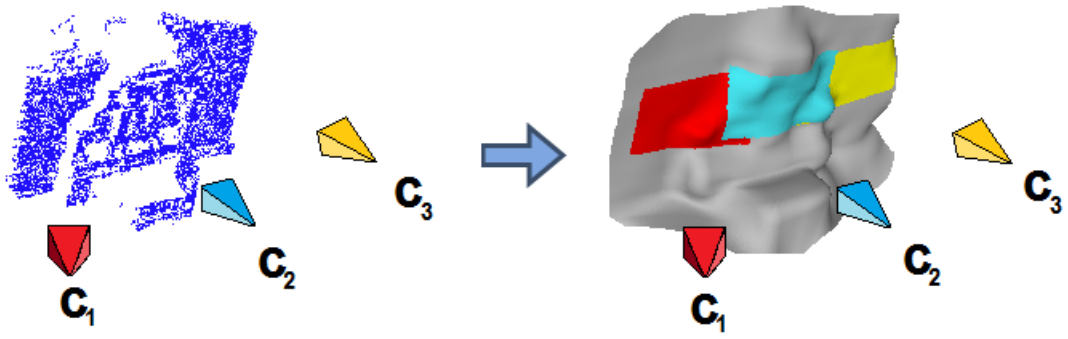


Figura 3.7: Proyección de vistas desde distintas poses de cámara, sobre superficie aproximada a partir de nube de puntos reconstruida.

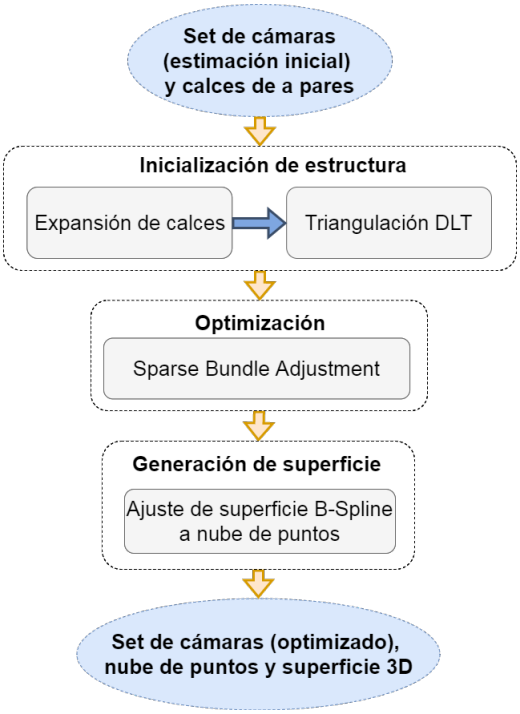


Figura 3.8: Diagrama de flujo de la reconstrucción de escena a partir de la estimación del conjunto de cámaras y los puntos de interés calzados.

Inicialización de estructura

La estructura inicial es generada a partir del método directo de triangulación por DLT sobre múltiples vistas descrito en [59], para lo cual se realiza previamente una expansión de los calces de a pares, la cual consiste en hacer un seguimiento secuencial de los puntos calzados a lo largo de toda la secuencia y almacenar la mayor cantidad de proyecciones conocidas para un mismo punto de la escena. Esta secuencia de puntos correspondientes (*point track*), es tomada como entrada para el algoritmo de triangulación, junto a las matrices de proyección de las cámaras, permitiendo obtener un punto $\mathbf{X}$ de la escena para cada *point track*, encontrando una solución *least-square* para el sistema de la Ec. 3.13, donde $\mathbf{p}_{\mathbf{j}}^{\mathbf{i}T}$ es el vector fila de la

matriz de proyección P_j de la vista j, así, cada proyección del punto aporta dos ecuaciones, requiriendo un mínimo de dos para poder resolver el sistema. La solución para el punto X, en coordenadas homogéneas, finalmente corresponde al vector propio de A^TA, asociado al menor valor propio de la matriz.

$$\begin{aligned} \begin{bmatrix} A_0 \\ \vdots \\ A_M \end{bmatrix} \mathbf{X} &= A\mathbf{X} = 0 \\ A_{\mathrm{i}} &= \begin{bmatrix} x_{\mathrm{i}}\mathbf{p_3^i}^T - \mathbf{p_1^i}^T \\ y_{\mathrm{i}}\mathbf{p_3^i}^T - \mathbf{p_2^i}^T \end{bmatrix} \end{aligned} \tag{3.13}$$

Optimización simultánea de poses de cámara y estructura

Según lo estudiado, prácticamente todos los trabajos que realizan reconstrucción de escena, poseen una etapa de ajuste final mediante *Bundle Adjustment*. El problema de optimización para el error cuadrático total de reproyección de los puntos de la escena, es definido en la Ec. 3.14. En esta ecuación, X_j representa el punto j de la escena, $x_{\mathrm{i}}j$ es el punto detectado en la imagen i, correspondiente a X_j, y $P_{\mathrm{i}}(X_j)$ corresponde al punto obtenido al proyectar X_j a través del modelo de la cámara asociada a la vista i.

$$\min_{\{\mathrm{P_i}\},\{\mathbf{X}_j\}} \sum_{\mathrm{i}=1}^{m} \sum_{j=1}^{n} ||\mathbf{x}_{\mathrm{i}j} - \mathrm{P_i}\mathbf{X}_j||^2 \tag{3.14}$$

Este problema de optimización puede ser realizado eficientemente si se explota el hecho de que las matrices involucradas son *sparse*, esto significa que, a pesar de tener altas dimensiones, pueden ser descompuestas por bloques matriciales de menor dimensión puesto que están estructuradas con grandes sectores que poseen valores nulos. El algoritmo de stitching propuesto considera la implementación del *Sparse Bundle Adjustment* presentada en [72].

Generación de superficie 3D

Hasta esta etapa, la reconstrucción obtenida es una nube de puntos de baja densidad, pero que cuenta con información importante sobre la estructura de la escena. Esto permite definir una superficie que represente una versión simplificada de la escena mediante el método utilizado en [66], donde se ajusta un modelo de superficie B-Spline a la nube de puntos.

En base a esto, se utiliza la propuesta de [71] que ajusta superficies *Non-Uniform Rational B-Splines* (NURBS), que consisten en una serie de curvas B-spline definidas por puntos de control y pesos que ponderan la influencia de los mismos sobre las curvas. Estos puntos se encuentran de manera ordenada, definiendo vértices en una malla de polígonos, permitiendo, a partir de una superficie inicial, agregar puntos de control desde la nube de puntos de manera iterativa, lo cual da forma a la superficie NURBS, como se muestra en la Fig. 3.10.

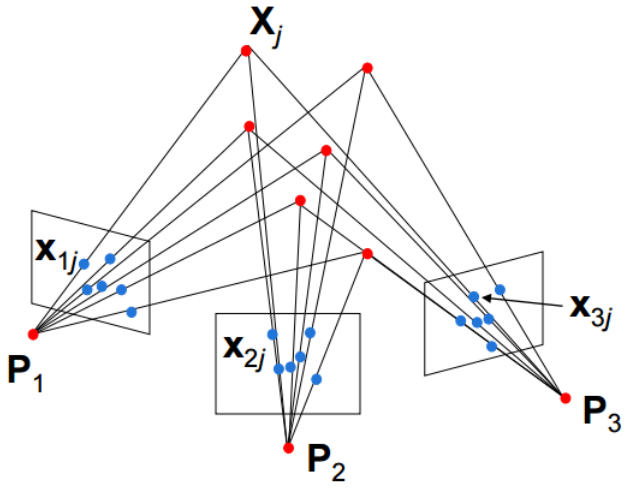


Figura 3.9: Reproyección de puntos de la escena

Posteriormente, se refinan iterativamente los pesos de los puntos de control sobre las curvas, manteniéndolos fijos y ajustando la superficie al resto de los puntos de la nube, minimizando la distancia entre cada punto y la superficie.

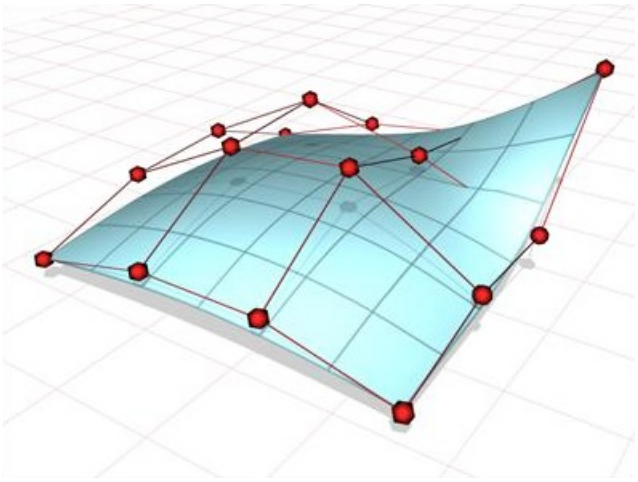


Figura 3.10: Modelo de superficie NURBS con puntos de control que definen su forma. [74]

Composición final sobre superficie 3D

Para finalizar la reconstrucción 3D de la escena, la proyección de cada imagen sobre la superficie se realiza de la misma forma que se hace en [66], utilizando imágenes de rango, obtenidas desde cada pose estimada de la cámara. Cada imagen de rango es construida a partir de la intersección de la superficie aproximada y de los haces proyectados desde el centro de la cámara hacia el espacio. Así, de existir la intersección, se obtiene un punto para cada par ordenado $(\mathrm{i}, j, \vec{r})$, donde (i, j) indica las coordenadas de la imagen y $\vec{r}$ representa la posición del punto de superficie más cercano proyectado en la imagen, expresado en las coordenadas de la cámara. Si no existe la intersección, el punto $\vec{r}$, será marcado como indefinido. Volviendo a

observar la Fig. 3.7, las imágenes coloreadas sobre la superficie corresponderían a las imágenes de rango de las distintas cámaras, definidas sobre la superficie mostrada.

En la práctica, para obtener la imagen de rango, en lugar de proyectar cada pixel sobre la superficie, se proyectan los vértices de una malla discreta que representa la superficie NURBS, hacia el plano de la imagen, almacenando su ubicación en el espacio y creando directamente una imagen de rango, cuya resolución dependerá de la discretización de la malla y los parámetros de la cámara usada para proyectar, permitiendo también interpolar algunos valores de rango indefinidos debido a la discretización. Este proceso tiene como resultado una nube de puntos coloreada para cada punto de la imagen que tiene proyección sobre la superficie, desde la vista respectiva. En la implementación del algoritmo, sólo se proyectan de manera independiente las imágenes en la superficie, creando nubes de puntos separadas para cada vista, pero bajo un mismo sistema de referencia. Así, el resultado de la composición puede ser observado mediante un software de visualización 3D como *MeshLab* [73], que permite cargar distintas nubes de puntos que, al estar en el mismo sistema de referencia, muestran el contenido combinado de todas las imágenes, pudiendo ser proyectado a cámaras con puntos de vista arbitrarios.

3.1.8. Adaptación para secuencias de vehículo en movimiento

Para aplicar la solución planteada al caso de una cámara estática capturando imágenes de un objeto en movimiento, en particular, sobre las imágenes del camión capturadas por el *Logmeter*, se considera el problema dual correspondiente al movimiento relativo de la cámara respecto a un objeto estático, como se propone en [66].

En el algoritmo desarrollado, la adaptación consiste principalmente en agregar una etapa previa de segmentación capaz de aislar del fondo la zona de la imagen que contiene al objeto en movimiento y, adicionalmente, filtrar las correspondencias establecidas entre pares de imágenes para descartar aquellas que pertenecen al fondo. Así, al diseño global modificado se presenta en la Fig. 3.11, donde se destacan en color rojo los bloques agregados respecto a la Fig. 3.1.

Segmentación del camión

Se utiliza el método de *frame difference* [21] sobre pares de imágenes consecutivas I_k e I_{k+1} para obtener una imagen binaria que identifique los pixeles en movimiento al fijar un umbral de binarización μ_{bin} adecuado sobre la imagen D_k obtenida al restar los cuadros, como se muestra en la Ec. 3.15. Debido a que las imágenes de movimiento suelen presentar ruido, se realiza un análisis de conectividad de regiones sobre la imagen binaria, como el planteado en la Sección 3.3.4 de [5], permitiendo eliminar aquellas que posean un área menor a un umbral definido previamente y, con el mismo criterio, eliminar los agujeros en la máscara correspondiente al área del camión al analizar las regiones conexas de la imagen complemento.

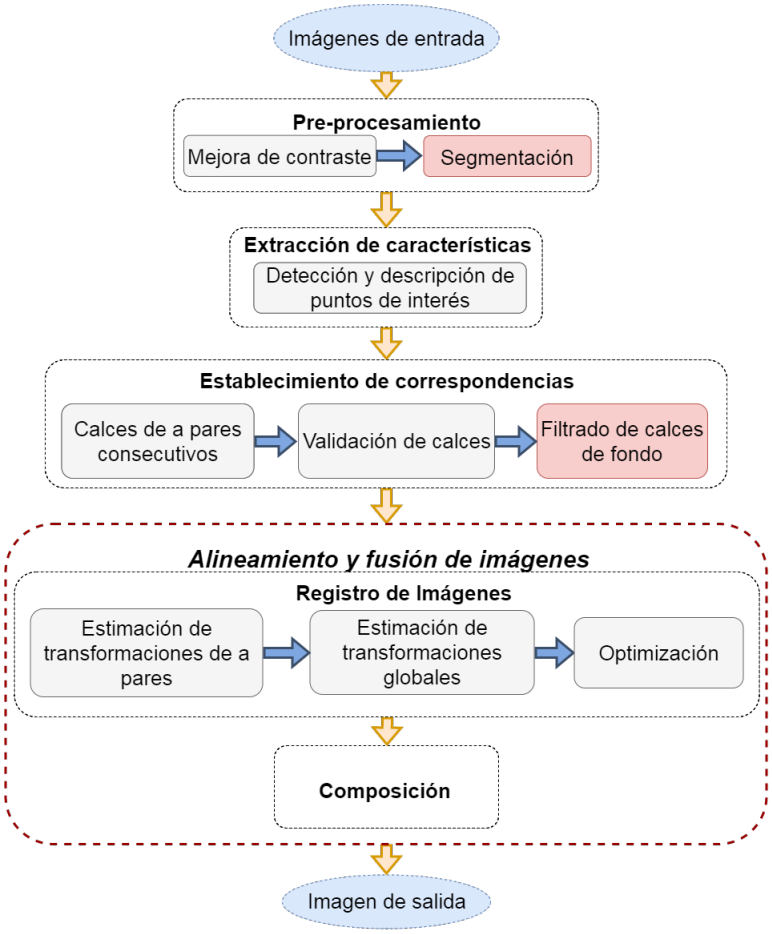


Figura 3.11: Diagrama global del algoritmo de stitching adaptado a presencia de objetos en movimiento.

$$\begin{aligned} D_k &= |I_k - I_{k+1}| \\ \text{Mask}_k &= D_k >= \mu_{bin} \end{aligned} \tag{3.15}$$

Cabe destacar que, dado que posteriormente se considera un filtrado de calces y la etapa de registro se realiza de manera robusta, no es necesario obtener una máscara perfecta. Sin embargo, se trata de reducir la cantidad de fondo residual de manera de poder descartar fácilmente los puntos de interés detectados en la zona del fondo y que la imagen compuesta contenga mayoritariamente el contenido alineado del camión. En el nivel de desarrollo planteado en este trabajo, esta etapa no es tan relevante para el registro de imágenes como el filtrado de calces, por lo que se deja propuesta la elección de un mejor método de segmentación para etapas futuras de desarrollo.

Filtrado de calces de fondo

La adaptación de imágenes resultantes que poseen objetos en movimiento requiere eliminar todo el contenido asociado al fondo del conjunto de correspondencias obtenido. Para ello se realiza una eliminación directa de los calces de puntos correspondientes al fondo, los cuales son identificados tomando como supuesto que los puntos del fondo son estáticos, por lo que el desplazamiento de los mismos debiese ser nulo entre cuadros consecutivos. Como en las imágenes existe un error de muestreo y presentan un cierto nivel de ruido, se propone fijar un umbral pequeño ($\mu_{\mathrm{d}} = 3[px]$) para aceptar sólo las correspondencias cuyo desplazamiento asociado sea mayor al umbral.

$$||\mathbf{x} - \mathbf{x'}|| \geq \mu_{\mathrm{d}} \tag{3.16}$$

3.1.9. Métricas

En el caso del registro de imágenes mediante homografías es posible medir la calidad del mismo en base a la calidad de la estimación de homografías. En este trabajo, la calidad de una homografía estimada sobre un par de imágenes, dado un conjunto de correspondencias validadas, es medida en base al desplazamiento promedio en pixeles de la reproyección a través de la homografía, lo que equivale a la distancia geométrica definida en la Ec. 3.17.

Además, al trabajar con secuencias ordenadas de imágenes, el registro final involucra la concatenación de las homografías estimadas de a pares hacia una imagen de referencia. Considerando como referencia la primera imagen, es posible obtener una cota inferior del desplazamiento existente en la proyección final, mediante la acumulación del desplazamiento medio definido. Con esto, se permite una evaluación que indicaría el peor escenario para la proyección de los puntos validados por homografías.

$$\langle \mathrm{d}(\mathrm{H}, \{\mathbf{x}, \mathbf{x}'\}_{\mathrm{i}=1\ldots n})\rangle = \frac{\sum_{\mathrm{i}=1}^{n} ||\mathbf{x}'_{\mathrm{i}} - \mathrm{H} \cdot \mathbf{x}_{\mathrm{i}}||}{n} \tag{3.17}$$

Por otro lado, el algoritmo de stitching por reconstrucción de escenas permite medir directamente el error total de reproyección cuadrático de todos los puntos reconstruidos, considerando todas las vistas en las que se proyectan, mediante la ecuación descrita por la Ec. 3.18.

$$\mathrm{e}\left(\{\mathrm{P}_{\mathrm{i}}\}_{\mathrm{i}=1\ldots m}, \{\mathbf{X}_j\}_{j=1\ldots n}\right) = \sum_{\mathrm{i}=1}^{m}\sum_{j=1}^{n} ||\mathbf{x}_{\mathrm{i}j} - \mathrm{P}_{\mathrm{i}}\mathbf{X}_j||^2 \tag{3.18}$$

En base a esta medida, se puede obtener una medida de error de reproyección en pixeles, equivalente al desplazamiento promedio en las distintas vistas, para todos los puntos de

la escena. Para ello, basta normalizar el cálculo del error total mediante la Ec. 3.19, que considera m vistas y n puntos.

$$\langle \mathrm{e}\left(\{\mathrm{P_i}\}_{\mathrm{i}=1\ldots m}, \{\mathbf{X}_j\}_{j=1\ldots n}\right)\rangle = \frac{\sum_{\mathrm{i}=1}^{m}\sum_{j=1}^{n} ||\mathbf{x}_{\mathrm{i}j} - \mathrm{P_i}\mathbf{X}_j||}{n \cdot m} \tag{3.19}$$

3.2. Plan de trabajo

La planificación del trabajo, enmarcada en un período de 15 semanas, correspondiente a la duración de un semestre regular, se expone en la carta Gantt de la Fig. 3.12. Este plan de trabajo, considera un período inicial para planificar el desarrollo del software y los períodos necesarios para desarrollar el prototipo inicial para la exploración del stitching mediante homografías y la versión posterior del algoritmo de stitching por reconstrucción de escena. Además, se considera el tiempo requerido para la evaluación y análisis de los resultados.

ACTIVIDAD	INICIO	DURACIÓN	SEMANAS 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Planificación de desarrollo	1	1															
Prototipo stitching por homografías	2	5															
Pre-procesamiento y segmentación	2	1															
Extracción de características	3	1															
Registro de imágenes	4	2															
Composición de imágenes	6	1															
Decisión de método de registro	7	-															
Versión stitching por reconstrucción escenas	7	7															
Triangulación	7	1															
Reconstrucción de escena	9	4															
Generación de superficie	12	1															
Composición de imagen	13	1															
Evaluación y análisis de resultados	14	2															

Figura 3.12: Carta Gantt de la planificación del trabajo.

3.2.1. Selección de bases de datos

En primer lugar, considerando que el trabajo planteado involucra secuencias de imágenes caracterizadas por un movimiento de captura donde existe una amplia traslación de la cámara, es necesario seleccionar secuencias que permitan evaluar las etapas tempranas del desarrollo y el resultado del mismo. Con este objetivo, se extraen secuencias de imágenes desde bases de datos comúnmente utilizadas para la evaluación de algoritmos de *Multi-view Stereo*, como las presentadas en [60] y [61], ya que, además de la secuencia, presentan información de los parámetros intrínsecos de la cámara utilizada, lo cual es requerido para el algortimo de stitching por reconstrucción de escenas. Debido a que el trabajo realizado introduce una propuesta para resolver el problema, en esta etapa de desarrollo no se hace necesario evaluar un número grande de secuencias con características similares, si no que se requiere probar el desempeño del algoritmo propuesto en distintos escenarios.

Por otro lado, tomando en cuenta la motivación que da pie a este proyecto, se seleccionarán secuencias de las capturas realizadas por el *Logmeter*, permitiendo evaluar el desempeño del algoritmo y sus etapas sobre casos de prueba en un escenario real de una posible aplicación del algoritmo, a lo largo de distintas etapas de desarrollo. Finalmente, para validar las etapas tempranas de la fase de exploración, se selecciona una secuencia sencilla de dos imágenes de un paisaje lejano, con la cámara en un punto fijo, permitiendo observar el desempeño de la propuesta para el problema clásico de composición de un panorama.

Capítulo 4

Resultados y Discusión

En este capítulo se presentan, en primera instancia, las bases de datos escogidas para probar el desarrollo realizado. Posteriormente, se definen las pruebas a realizar y, luego, se exponen los resultados obtenidos, junto a sus respectivas discusiones. Por último, se finaliza el capítulo con una discusión general del trabajo desarrollado.

4.1. Bases de datos

Para el desarrollo del algoritmo de stitching, se deben escoger bases de datos adecuadas, que contengan la información necesaria de la captura y, además, permitan probar el desarrollo de cada etapa y del resultado final. A continuación, se presentan las secuencias de imágenes escogidas en base a los trabajos estudiados, mostrando ejemplos de las imágenes contenidas que resumen el rango de movimiento total en la secuencia.

Considerando la motivación del proyecto, entre las bases de datos utilizadas, se consideran secuencias de imágenes extraídas de los archivos de captura del *Logmeter*, proporcionados por Woodtech S.A.. En el Anexo 6.1, se muestran las imágenes de las secuencias completas y se detallan aspectos técnicos de las mismas.

1. **'Temple'**:
   Corresponde a una secuencia de 12 imágenes, equivalentes al intervalo $[13, 24]$ de la base de datos *TempleRing* [60], donde el movimiento de la cámara describe una circunferencia alrededor de una maqueta del templo de los Dioscuros de Agrigento.

Figura 4.1: Ejemplo de imágenes de la secuencia 'Temple'.

2. **'Fountain'**: Equivale a la secuencia `fountain-P11` [61], donde la escena capturada corresponde a un muro con una fuente de agua.

Figura 4.2: Ejemplo de imágenes de la secuencia 'Fountain'.

3. **'HerzJesu'**: Corresponde a las imágenes del intervalo $[0, 13]$ de la secuencia `HerzJesu-P25` [61], que muestran la entrada de un edificio religioso con arquitectura compleja.

Figura 4.3: Ejemplo de imágenes de la secuencia 'HerzJesu'.

4. **'SceauxCastle'**: Base de datos que captura el *Château de Sceaux* en una secuencia de 11 imágenes, junto a la matriz de calibración intrínseca de la cámara utilizada, disponible en [75].

Figura 4.4: Ejemplo de imágenes de la secuencia 'SceauxCastle'.

5. **'PaisajeNavarino'**: Secuencia de dos imágenes, creada con el propósito de validar el desarrollo realizado para el algoritmo de stitching por homografías sobre una escena estática y sin traslaciones amplias de cámara. Corresponde a dos imágenes de un paisaje en la Isla Navarino, capturadas con la cámara de un *smartphone*, rotando levemente el dispositivo.

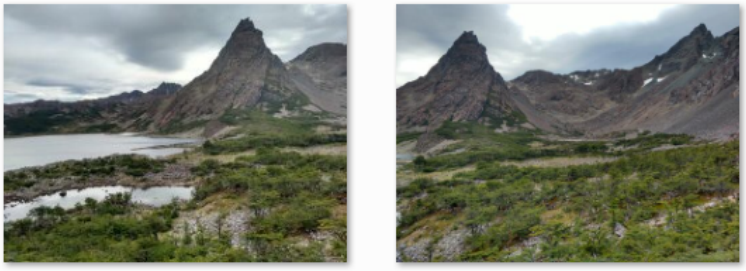


Figura 4.5: Secuencia de imágenes 'PaisajeNavarino'.

Capturas del *Logmeter*

Se consideran, por separado, cuatro secuencias de imágenes, capturadas sobre dos camiones con cargamento de madera desde las cámaras estáticas laterales del *Logmeter*, a una tasa de 1[fps]. Cada secuencia tiene un número diferente de imágenes. Estas son seleccionadas escogiendo manualmente las imágenes que contienen el camión y su cargamento, descartando aquellas imágenes donde no aparece aún el camión.

6. **'Truck1R'**: Incluye 21 imágenes de la vista lateral derecha del primer camión.

Figura 4.6: Ejemplo de imágenes de la secuencia 'Truck1R'.

7. **'Truck1L'**: Incluye 18 imágenes de la vista lateral izquierda del primer camión.

Figura 4.7: Ejemplo de imágenes de la secuencia 'Truck1L'.

8. **‘Truck2R’**: Incluye 14 imágenes de la vista lateral derecha del segundo camión.

Figura 4.8: Ejemplo de imágenes de la secuencia ‘Truck2R’.

9. **‘Truck2L’**: Incluye 19 imágenes de la vista lateral izquierda del segundo camión.

Figura 4.9: Ejemplo de imágenes de la secuencia ‘Truck2L’.

4.2. Definición de pruebas

La primera fase de desarrollo planificada, corresponde a un prototipo que realiza stitching por registro de homografías globales. En este prototipo, se definen e implementan las etapas de pre-procesamiento, extracción de características y establecimiento de correspondencias, además de las etapas propias del alineamiento y composición de imágenes. Para evaluar el desarrollo inicial y las decisiones tomadas en el diseño, se definen las siguientes pruebas sobre el prototipo desarrollado.

- Evaluación de la mejora de contraste sobre la extracción de características y cantidad de correspondencias establecidas.
- Prueba del prototipo de stitching por homografías sobre secuencia ‘PaisajeNavarino’.
- Medición del desplazamiento medio de reproyección de a pares y desplazamiento medio acumulado hacia primera imagen.
- Evaluación visual del resultado de stitching por homografías sobre secuencias representativas.

Por otro lado, en la segunda fase de desarrollo, se implementa el stitching por reconstrucción de escena, para lo cual se definen las pruebas enumeradas a continuación.

- Medición del error de reproyección promedio de estructura inicial y final, luego de la optimización.
- Evaluación visual del resultado de nube de puntos de la estructura, superficie ajustada y composición de imágenes en modelo 3D.

4.3. Resultados y discusión

Para desarrollar un programa capaz de ejecutar el algoritmo propuesto, dentro del tiempo estipulado, se hace uso de las funcionalidades de las librerías OpenCV, PCL y SBA, además de las propias librerías internas de la empresa Woodtech S.A.. Así, utilizando el programa desarrollado, se realizan las pruebas definidas sobre las secuencias escogidas, permitiendo evaluar la validez conceptual del diseño propuesto e identificar los puntos claves de mejora. Los resultados son presentados en cuatro secciones: adaptación del algoritmo para secuencias de vehículo en movimiento, evaluación de la mejora de contraste sobre la extracción de características y cantidad de correspondencias establecidas, prototipo de algoritmo de stitching por homografías globales y algoritmo de stitching por reconstrucción de escena. Cada una de estas secciones contiene la respectiva discusión sobre los aspectos más relevantes, tanto del diseño como de los resultados obtenidos en las etapas involucradas.

4.3.1. Adaptación para secuencias de vehículo en movimiento

A continuación, se presentan resultados que ejemplifican la adaptación del algoritmo propuesto a las secuencias de imágenes que contienen un vehículo en movimiento, capturadas en el *Logmeter*, junto a la discusión pertinente. Para visualizar el funcionamiento de las etapas de segmentación y filtrado de calces de fondo, se utilizan como referencia los dos pares de imágenes mostrados en la Fig. 4.10, correspondientes al primer y último par de imágenes consecutivas de la secuencia 'Truck1R', luego de la mejora de contraste (ver Anexo 6.1).

(a) Primer par de imágenes consecutivas. (b) Último par de imágenes consecutivas.

Figura 4.10: Pares de imágenes de referencia, para ejemplificar segmentación y filtrado de calces de fondo en secuencia 'Truck1R'.

Los ejemplos de las Fig. 4.11 y 4.12, muestran resultados de la segmentación por *frame-difference* y del procesamiento de la máscara por análisis de conectividad, con un umbral de binarización $\mu_{\text{bin}} = 33$ y un tamaño mínimo de región de movimiento correspondiente al 0,6 % del área abarcada en la imagen de entrada.

Utilizando los mismos ejemplos, con el objetivo de visualizar los resultados de la etapa de filtrado de calces de fondo por umbral de desplazamiento mínimo, se tomaron las imágenes y calces obtenidos sin realizar la etapa de segmentación. Sin embargo, en el algoritmo planteado, esta sí es incluida. El umbral de desplazamiento mínimo utilizado es fijado en $\mu_{\text{d}} = 3[px]$, logrando los resultados mostrados en las Fig. 4.13 y 4.14, donde se muestran los calces como segmentos verdes que unen los puntos de interés respectivos de cada imagen.

(a) Máscara inicial. (b) Máscara final. (c) Imagen segmentada.

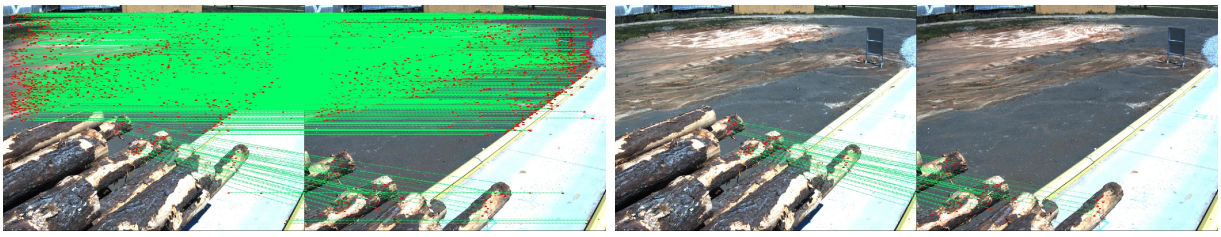
Figura 4.11: Ejemplo de segmentación por movimiento para la primera imagen de la secuencia 'Truck1R'.

(a) Máscara inicial. (b) Máscara final. (c) Imagen segmentada.

Figura 4.12: Ejemplo de segmentación por movimiento para la penúltima imagen de la secuencia 'Truck1R'.

(a) Calces de puntos de interés iniciales. (b) Calces de puntos de interés filtrados.

Figura 4.13: Ejemplo de filtrado de calces de fondo por desplazamiento mínimo en el primer par de imágenes de la secuencia 'Truck1R'.



(a) Calces de puntos de interés iniciales. (b) Calces de puntos de interés filtrados.

Figura 4.14: Ejemplo de filtrado de calces de fondo por desplazamiento mínimo en el último par de imágenes de la secuencia 'Truck1R'.

Discusión sobre etapas de adaptación

En primer lugar, se identifica que la etapa de filtrado de calces de fondo en base a un desplazamiento mínimo cumple su función, permitiendo incluso prescindir de la etapa de segmentación previa a la extracción de características. Tanto la Fig. 4.13, como la Fig. 4.14, ejemplifican la eliminación por completo de los calces del fondo en la secuencia con el camión en movimiento, permitiendo que la etapa de registro que le sigue, refleje los cambios inducidos por un movimiento relativo de la cámara respecto al camión. Un detalle importante, es que los calces que sirven de entrada a esta etapa deben haber pasado una validación previa, comprobando que se mantiene una relación geométrica global entre los puntos correspondientes. En el caso de existir correspondencias incorrectas, es probable que el umbral de desplazamiento mínimo sea fácilmente superado, por lo que, en el caso de corresponder al fondo, no se podría filtrar bajo este criterio.

En segundo lugar, respecto a la etapa de segmentación incorporada en el diseño, es necesario destacar que su utilidad, en las etapas iniciales, consiste únicamente en reducir el espacio donde se extraen características y se buscan correspondencias, ya que el resto de las etapas utilizan directamente las características extraídas y los calces establecidos, por lo que la imagen misma pierde relevancia. En este sentido, se puede observar en la Fig. 4.11, que no se logra una segmentación precisa debido a la simplicidad de la técnica abordada, pero sí es posible lograr el objetivo de acotar el espacio de búsqueda de características. En otro aspecto, para las etapas de proyección y composición, debido a que el contenido de las imágenes es alineado en base a la zona del camión en movimiento, el fondo presente en las imágenes, al ser proyectado mediante homografías, posee grandes distorsiones de perspectiva, afectando la visualización de la composición final. Así, un punto de mejora para la visualización sería la etapa de segmentación, ya que el método escogido de segmentación por *frame-difference* no elimina por completo el fondo.

4.3.2. Evaluación de la mejora de contraste sobre la extracción de características y cantidad de correspondencias establecidas

En esta etapa, se evaluó el efecto de la mejora de contraste sobre la cantidad de puntos de interés encontrados con el extractor AKAZE sobre las distintas secuencias, con y sin mejora de contraste. Los resultados se encuentran condensados en la Tabla 4.1, que muestra el porcentaje de aumento al utilizar CLAHE, relativo a la cantidad de puntos de interés detectados sin su uso.

En el mismo contexto, para apreciar el efecto de la mejora de contraste sobre la cantidad de correspondencias establecidas se compara el número total de calces validados, con y sin la aplicación de CLAHE, utilizando la validación de calces por homografía global para descartar las correspondencias incorrectas. La Fig. 4.15 resume los resultados obtenidos, mostrando una clara mejora en la cantidad de correspondencias establecidas.

Secuencia	Aumento Promedio [%]
'Temple'	43,96
'Fountain'	818,24
'HerzJesu'	345,20
'SceauxCastle'	271,21
'Truck1R'	65,34
'Truck1L'	26,35
'Truck2R'	79,62
'Truck2L'	54,32

Tabla 4.1: Aumento porcentual promedio de cantidad de puntos de interés detectados para las distintas secuencias.

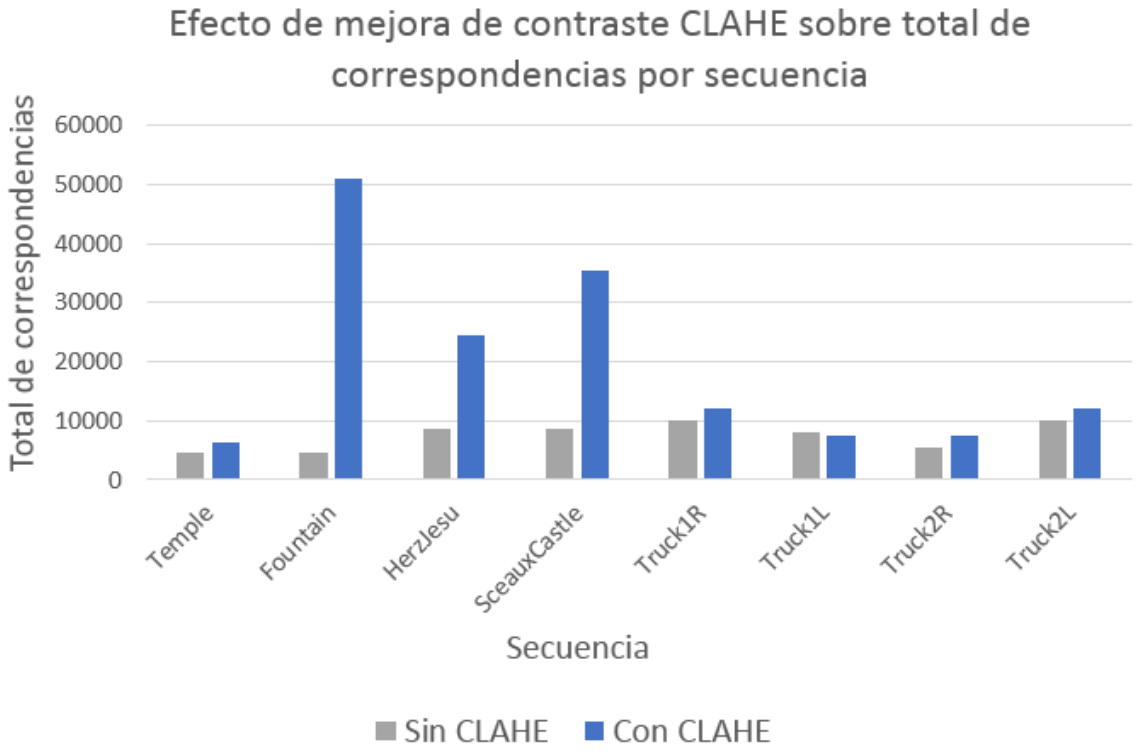


Figura 4.15: Efecto de mejora de contraste CLAHE sobre total de correspondencias validadas por secuencia.

Discusión sobre efectos de la mejora de contraste

Considerando los resultados desplegados en la Tabla 4.1, es directo observar que la cantidad de puntos de interés detectados mediante el extractor AKAZE, es potenciada al realizar un pre-procesamiento con la mejora de contraste CLAHE, logrando incluso en la secuencia 'Fountain' un aumento porcentual promedio de más de 800 %. Esto tiene la ventaja potencial de permitir encontrar correspondencias en zonas que no serían abarcadas de no realizarse la mejora de contraste, lo cual a su vez permite que las transformaciones estimadas representen mejor el cambio respectivo entre las imágenes correspondientes.

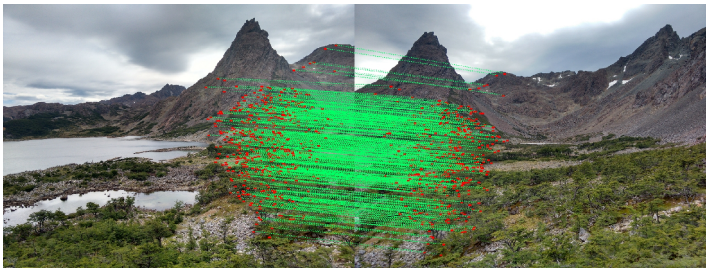
Por otro lado, evaluando la cantidad de correspondencias establecidas a partir de los puntos de interés detectados, al utilizar CLAHE, se observa un aumento considerable que, en el caso del algoritmo de stitching por reconstrucción de escena, se traduce directamente en un mayor número de puntos en la estructura parcial reconstruida. En las secuencias correspondientes a capturas del *Logmeter*, este efecto es mucho menor al de las secuencias 'Fountain', 'HerzJesu' y 'SceauxCastle'. Sin embargo, se considera que, dados los resultados obtenidos, la etapa de mejora de contraste debe ser incluida en el diseño final del algoritmo de stitching, sobretodo en su versión por reconstrucción de escena.

4.3.3. Prototipo de algoritmo de stitching por homografías globales

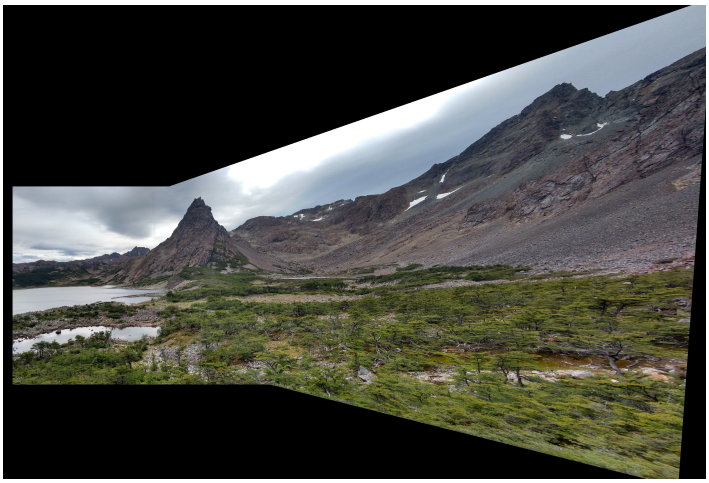
Mediante el prototipo implementado que ejecuta el algoritmo de stitching con registro por homografías globales, se realizaron las pruebas definidas en la Sección 4.2, permitiendo validar el desarrollo realizado en las etapas previas al registro de imágenes, además de observar las limitaciones de esta técnica cuando se utilizan transformaciones de homografía sobre escenas complejas que no se pueden aproximar por un plano. A continuación, se muestran los resultados de la prueba inicial sobre la secuencia 'PaisajeNavarino', para luego presentar los resultados de medición de desplazamiento medio y la evaluación visual. Finalmente, se discute en torno a los aspectos relevantes y la decisión tomada de continuar el desarrollo mediante un algoritmo de stitching por reconstrucción de escena.

Prueba inicial del algoritmo

Para validar la implementación del prototipo de stitching por homografías, se probó el resultado sobre el par de imágenes de la secuencia 'PaisajeNavarino', que presenta un paisaje donde la mayor parte del contenido es lejano y se aproxima bien a una escena plana, además, el movimiento de la cámara consiste en rotación pura. En la Fig. 4.16, se muestran las correspondencias validadas mediante la homografía estimada con un umbral de $\mu_{\text{RANSAC}} = 1{,}0[px]$ para RANSAC y la imagen obtenida como composición tomando como referencia el plano de la primera imagen y utilizando la superposición de la imagen proyectada en la zona de traslape.



(a) Correspondencias validadas por estimación de homografía global.



(b) Composición proyectando sobre el plano de la primera imagen.

Figura 4.16: Prueba inicial del prototipo de stitching por homografías, sobre secuencia 'Paisaje Navarino'.

Medición del desplazamiento medio de reproyección

Los resultados mostrados en las Fig. 4.17 y 4.18, corresponden a las mediciones realizadas sobre las secuencias 'Fountain' y 'Truck2L', respectivamente, aplicando un umbral de $\mu_{\text{RANSAC}} = 1{,}0[px]$ para la estimación por RANSAC, ya que se desea un alineamiento preciso. Además, esta prueba también fue realizada sobre las secuencias 'Temple' y 'Truck1R', cuyos resultados pueden ser encontrados en el Anexo 6.3.

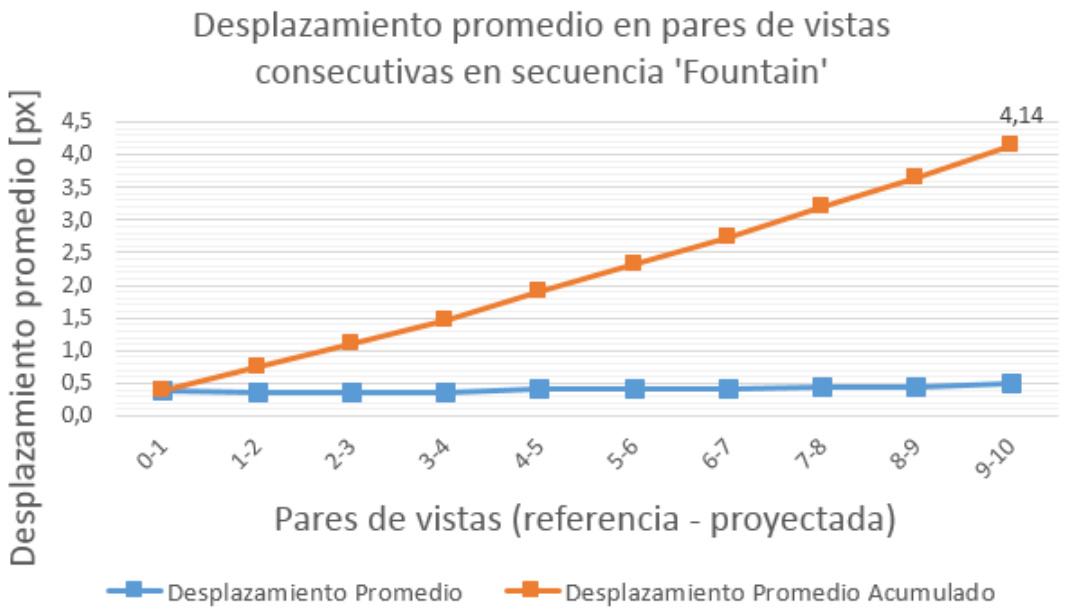


Figura 4.17: Desplazamiento promedio en pares de vistas consecutivas en secuencia 'Fountain'. El desplazamiento acumulado para llevar la última vista hacia la primera es de 4,14[px]

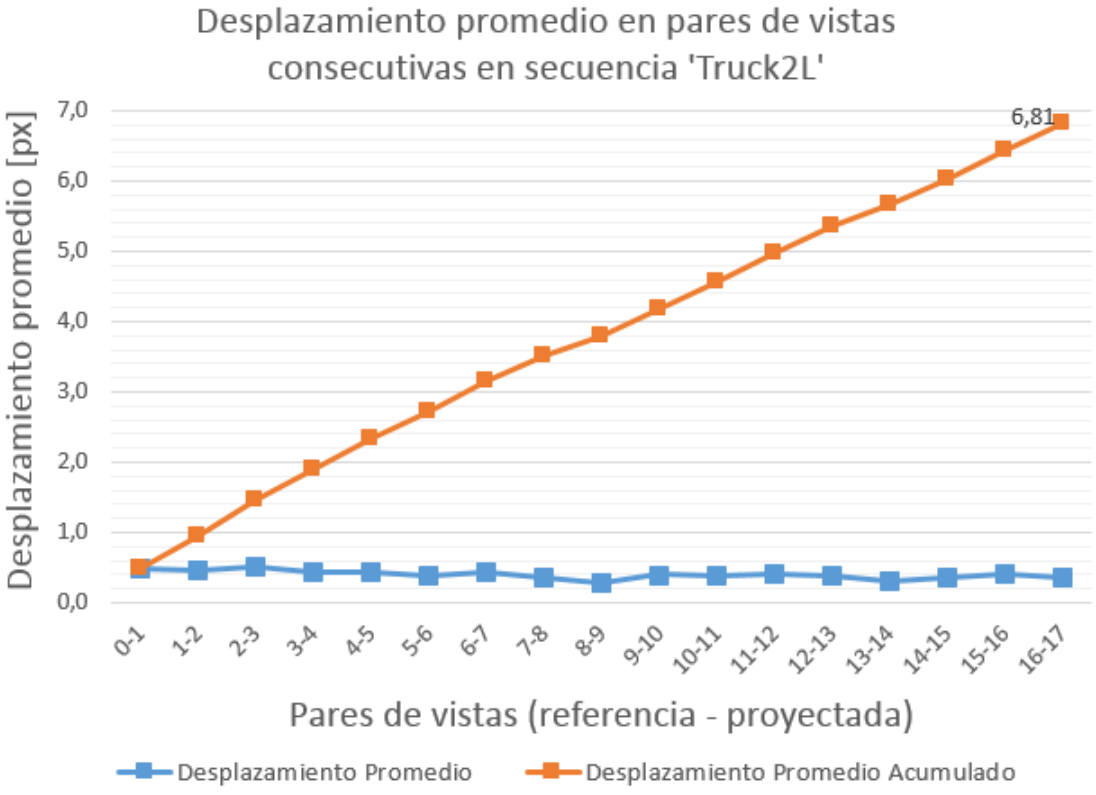


Figura 4.18: Desplazamiento promedio en pares de vistas consecutivas en secuencia 'Truck2L'. El desplazamiento acumulado para llevar la última vista hacia la primera es de 6,81[px]

Evaluación visual

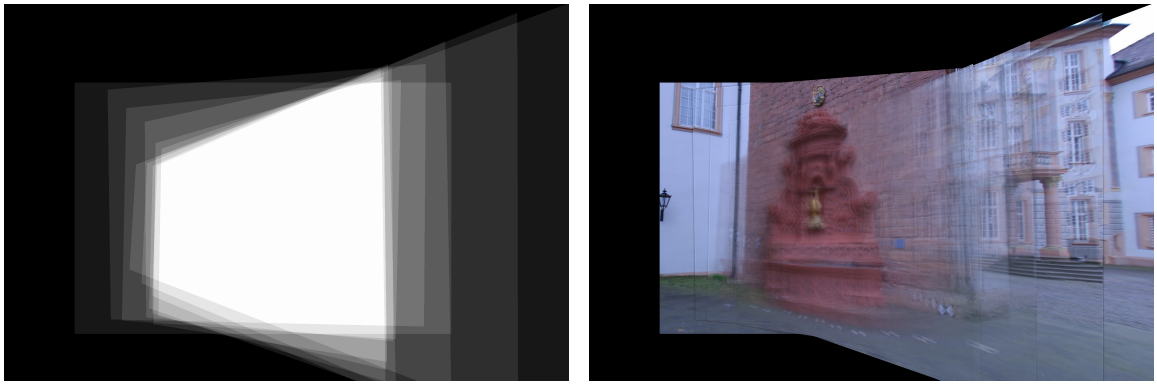
A continuación, se presentan los resultados finales del algoritmo de stitching que realiza un registro por homografías, sobre las secuencias 'Fountain' y 'Truck2L'. Para mostrar con mayor nivel de detalle los resultados de la secuencia 'Truck2L', se utiliza el segmento correspondiente al intervalo $[3, 9]$ de las vistas. Además, se ejemplifican los resultados de etapas claves, como el establecimiento de correspondencias válidas en base a imágenes representativas que facilitan el análisis respectivo. Finalmente, se presenta la composición obtenida aplicando las técnicas de fusión por superposición y *blending* lineal en las zonas de traslape.

Las ejecuciones del algoritmo utilizan los mismos parámetros de estimación definidos para la prueba anterior. En particular, para las secuencias del *Logmeter*, se aplican las etapas de segmentación y filtrado de calces de fondo con los parámetros fijados en la Sección 4.3.1. En el Anexo 6.3, están disponibles los resultados finales de la composición con *blending*, obtenidos para las secuencias 'Temple', 'Truck1R' y la secuencia 'Truck2L' completa.

(a) Proyección de máscaras. (b) Composición por *blending* lineal.

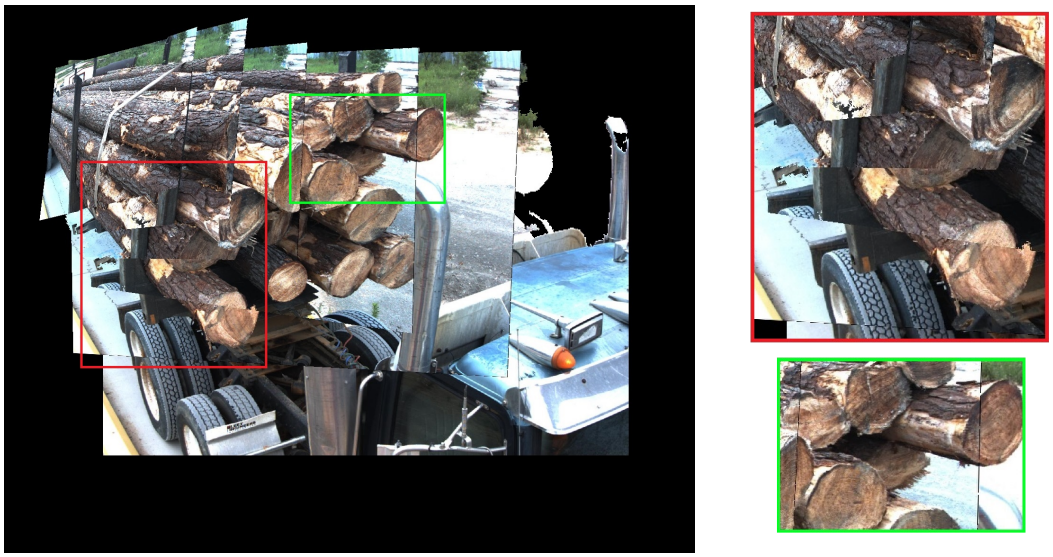
Figura 4.19: Resultado de la composición de la secuencia que abarca el intervalo $[3, 9]$ de 'Truck2L', mediante registro por homografías globales hacia el plano de referencia de la primera imagen.

En las Fig. 4.19 y 4.20, se muestra la proyección de las máscaras, permitiendo apreciar las zonas de traslape. Además, se observa la composición utilizando *blending* lineal para el contenido compartido en estas zonas. Por otro lado, las Fig. 4.21 y 4.22, presentan el resultado de la composición por superposición, donde se han amplificado zonas relevantes para el análisis. Aquellas zonas demarcadas con un contorno rojo, destacan regiones donde el contenido mezclado queda fuertemente desalineado, mientras que las zonas de contorno verde ejemplifican áreas que contienen una alineación local cercana a la correcta. Adicionalmente, se incluyen, a modo de ejemplo, imágenes representativas que muestran puntos de interés correspondientes a los calces validados por la homografía que relaciona la imagen siguiente de la secuencia con la imagen mostrada. Así, se intenta ilustrar cuáles son los planos relacionados por las homografías, sobre los cuales se proyecta al momento de realizar la composición.



(a) Proyección de máscaras. (b) Composición por *blending* lineal.

Figura 4.20: Resultado de la composición de la secuencia 'Fountain', mediante registro por homografías globales hacia primera imagen.

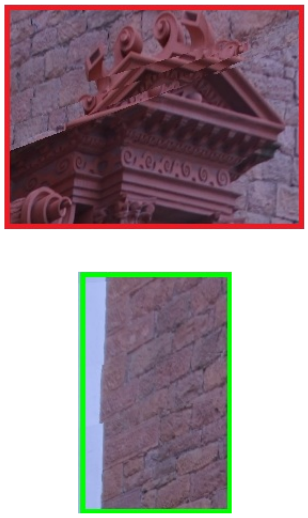


(a) Composición por superposición. (b) Zonas resaltadas.

(c) Imágenes representativas con puntos de interés correspondientes a los calces validados por la homografía con la imagen siguiente.

Figura 4.21: Resultado de stitching de la secuencia que abarca el intervalo $[3, 9]$ de 'Truck2L', mediante registro por homografías globales hacia primera imagen.

(a) Composición por superposición.



(b) Zonas resaltadas.

(c) Imágenes representativas con puntos de interés correspondientes a los calces validados por la homografía con la imagen siguiente.

Figura 4.22: Resultado de stitching de la secuencia 'Fountain', mediante registro por homografías globales hacia primera imagen.

Discusión del prototipo de stitching por homografías

Observando los resultados de la prueba inicial del algoritmo sobre la secuencia 'PaisajeNavarino', se puede apreciar en la Fig. 4.16, que las correspondencias establecidas y validadas al estimar la homografía global, relacionan de manera correcta los puntos de interés en la zona de cada imagen que posee contenido compartido. Esto conlleva a un alineamiento preciso que, al momento de proyectar y realizar la composición mostrada, deja un borde prácticamente imperceptible. Con este experimento, se cumple el objetivo de validar el correcto funcionamiento del sistema de stitching desarrollado, en una condición cercana a la ideal, como es el caso de la secuencia 'PaisajeNavarino', que tiene una rotación pura de la cámara y captura una escena lejana.

La siguiente fase, consiste en probar el prototipo sobre secuencias capturadas con una cámara con amplios movimientos de traslación. En esta prueba, se mide el desplazamiento promedio, en pixeles, entre pares de imágenes consecutivas para cada una de las secuencias. En los resultados presentados en las Fig. 4.17 y 4.18, es posible notar que, si bien el desplazamiento promedio obtenido entre pares consecutivos es menor a $1[px]$, al momento de concatenar las homografías desde la última imagen hacia la primera, se obtienen desplaza-

mientos promedio mayores a $4[px]$. Esto, considerando que la medición del desplazamiento promedio se realiza sólo a partir de los puntos que están en consenso con la homografía, representa un límite para la precisión de alineamiento que se puede obtener mediante esta técnica.

Por otro lado, las zonas correspondientes a los puntos que no son representados por las homografías estimadas de a pares, quedan completamente desalineadas luego de componer las transformaciones sucesivas, como se aprecia en las Fig. 4.21 y 4.22, donde las homografías se ajustan al plano de una cara frontal de un tronco o a un lado de ellos, en el caso de la secuencia 'Truck2L', y al plano de la muralla, en el caso de la secuencia 'Fountain'. Observando con mayor detalle las zonas amplificadas en ambas imágenes y los puntos de interés representados por homografías, se hace evidente que la estimación de una homografía global precisa para relacionar un par de imágenes consecutivas del tipo de secuencias abordadas, no es el enfoque adecuado.

Como fue mencionado en la Sección 3.1.4, en este punto del desarrollo se plantean dos opciones: continuar con el registro mediante homografías, estimando de manera local transformaciones para cada zona que se aproxime a un plano distinto, o bien, realizar un registro de las poses de las cámaras, obteniendo simultáneamente una reconstrucción parcial de la escena. En el primer enfoque, para una buena representación de las escenas complejas, se tendrían que estimar planos locales en áreas pequeñas de la imagen, lo que requeriría tener al menos cuatro correspondencias en cada zona, lo cual, según se estudió en el estado del arte, es una tarea muy compleja en este tipo de imágenes. Así, el segundo camino se presenta como una alternativa viable, que, además, permite obtener una reconstrucción parcial de la escena, lo cual es de alto interés para la empresa en la que se desarrolla el trabajo. Por este motivo, el diseño del algoritmo de stitching desarrollado contempla una etapa de alineamiento y composición basada en la reconstrucción de escena.

4.3.4. Algoritmo de stitching por reconstrucción de escena

A continuación, se presentan los resultados y discusión respecto al desarrollo y pruebas realizadas en el algoritmo de stitching por reconstrucción de escena.

Medición de error de reproyección promedio de la reconstrucción parcial

Contando con las correspondencias establecidas entre pares de imágenes consecutivas, la estimación de poses de cámara relativas es realizada mediante RANSAC con $\mu_{\text{RANSAC}} = 1{,}0[px]$, para luego concatenar las transformaciones euclidianas que las definen y, así, obtener las poses de cámara en un sistema de referencia común, fijado en las coordenadas de la cámara de la primera vista. Luego, expandiendo los calces a lo largo de todas las vistas, por medio de la triangulación se obtiene una reconstrucción parcial inicial aproximada, la cual es posteriormente optimizada mediante la etapa de *Bundle Adjustment*, llegando a una nube de puntos final.

De esta forma, al ejecutar la implementación del algoritmo de stitching por reconstrucción de escena, se obtienen los resultados resumidos en la Tabla 4.2, donde se muestran, para las distintas secuencias, la cantidad de puntos de escena reconstruidos, los errores de reproyección promedio inicial de la estructura triangulada y los errores de reproyección promedio de la nube de puntos final, luego de realizar la optimización.

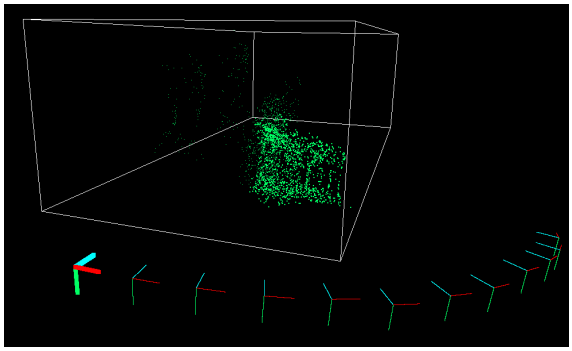
Secuencia	Cantidad de puntos reconstruidos	Error de reproyección promedio inicial [px]	Error de reproyección promedio final [px]
'Temple'	2435	1,1440	0,0059
'Fountain'	23942	0,4362	0,0017
'HerzJesu'	4528	26,6981	0,0045
'SceauxCastle'	6943	0,8416	0,0081
'Truck1R'	3774	1,2397	0,0046
'Truck1L'	3181	1,5103	0,0053
'Truck2R'	3094	1,3837	0,0056
'Truck2L'	4860	1,3135	0,0048

Tabla 4.2: Número de puntos reconstruidos y errores de reproyección promedio inicial y final para las secuencias evaluadas.

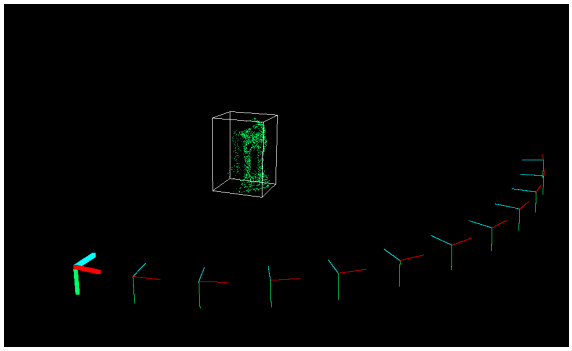
Evaluación visual

Los resultados de la evaluación visual de las etapas de reconstrucción de escena, son presentados mediante los ejemplos expuestos en las Fig. 4.23, 4.24 y 4.25, donde se muestran las nubes de puntos y conjunto de cámaras estimadas en las secuencias 'Temple', 'HerzJesu' y 'Fountain', respectivamente, antes y después del *Bundle Adjustment*, junto a la superficie ajustada. Las imágenes desplegadas, capturadas desde el software de visualización 3D *MeshLab*, corresponden al mismo punto de vista y muestran las cámaras representadas con los ejes del sistema de coordenadas de cada cámara, siendo el de color celeste el eje óptico, que

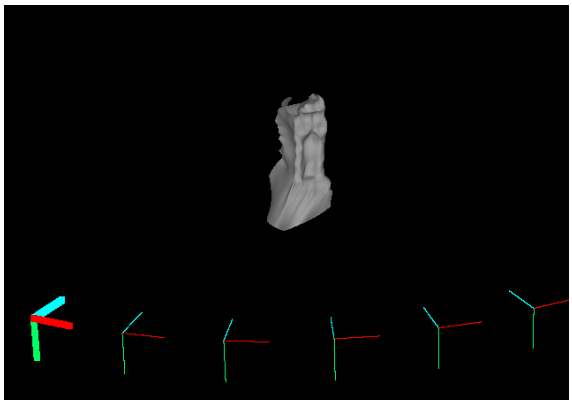
indica la dirección hacia la que apunta la cámara. La representación dibujada con un mayor grosor indica la primera cámara de la secuencia, utilizada como sistema de referencia global, tanto para los puntos como para las otras cámaras.



(a) Reconstrucción parcial inicial.

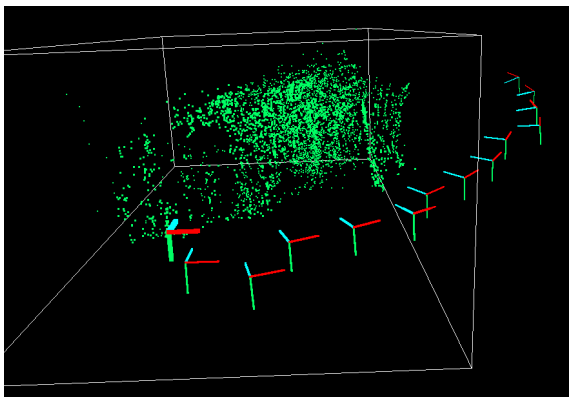


(b) Reconstrucción parcial final.

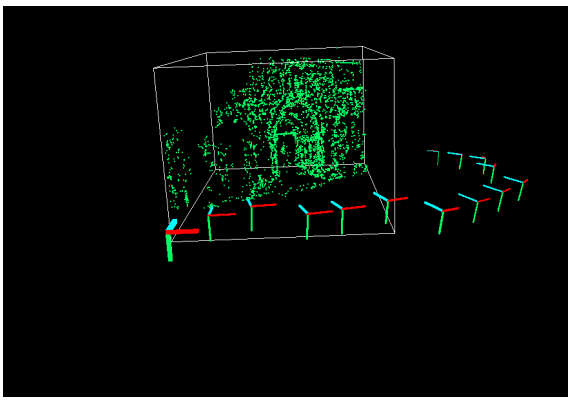


(c) Superficie ajustada para proyección.

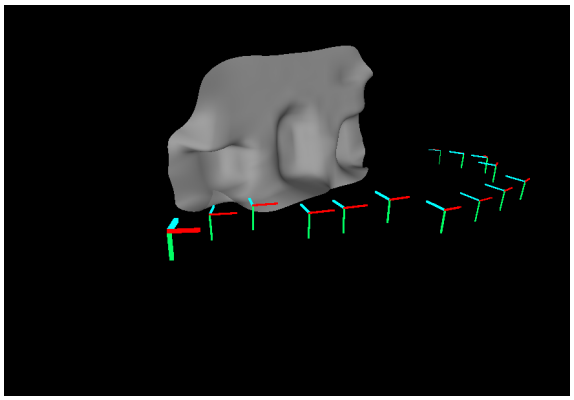
Figura 4.23: Resultado de la reconstrucción de escena sobre la secuencia 'Temple'.



(a) Reconstrucción parcial inicial.



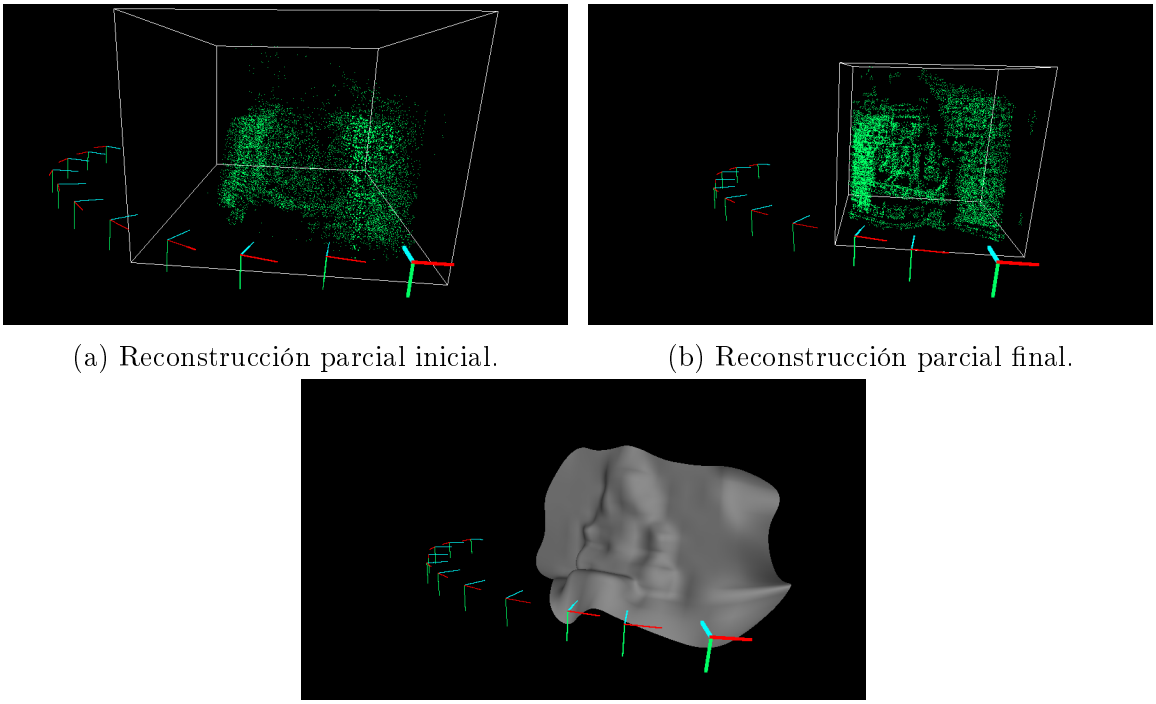
(b) Reconstrucción parcial final.



(c) Superficie ajustada para proyección.

Figura 4.24: Resultado de la reconstrucción de escena sobre la secuencia 'HerzJesu'.

Se presentan a continuación los resultados finales del algoritmo de stitching por reconstrucción de escena, siendo mostradas, en dos vistas, la nube de puntos final, la superficie ajustada y la composición obtenida para la secuencia 'Fountain' en la Fig. 4.26. En la Fig. 4.27, es posible observar las proyecciones de las imágenes por separado, facilitando el análisis posterior. Siguiendo una estructura similar, las Fig. 4.28 y 4.29, presentan los resultados para la secuencia 'Truck2L', mostrando las proyecciones progresivas sobre la composición final. En el Anexo 6.4 se pueden encontrar los resultados obtenidos a las otras secuencias.



(a) Reconstrucción parcial inicial.

(b) Reconstrucción parcial final.

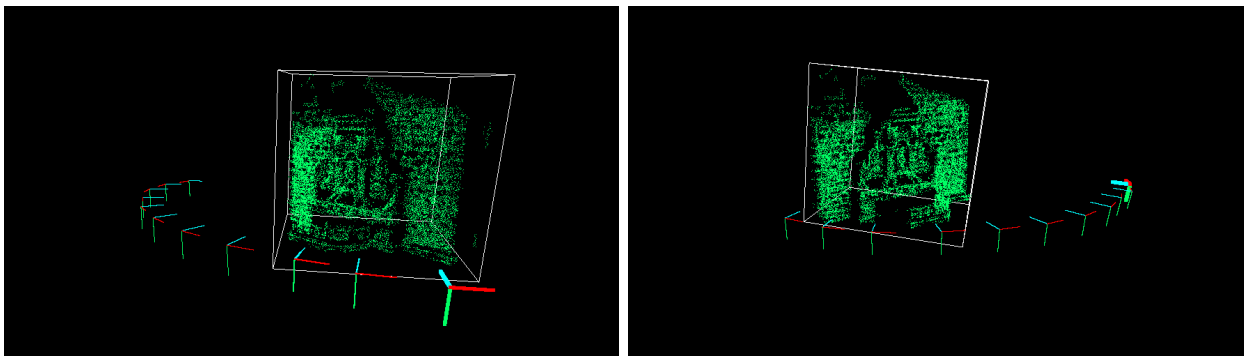
(c) Superficie ajustada para proyección.

Figura 4.25: Resultado de la reconstrucción de escena sobre la secuencia 'Fountain'.

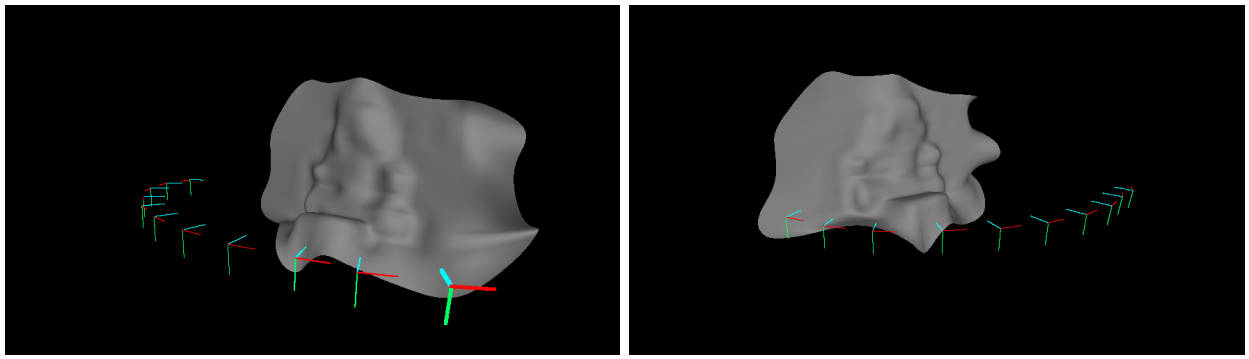
Discusión del desarrollo de stitching por reconstrucción de escena

Analizando los resultados presentados en la Tabla 4.2, se observa que la etapa de *Bundle Adjustment*, permite lograr un error de reproyección promedio de 0,005$[px]$, con desviación de 0,0018$[px]$, lo que se traduce en que la nube de puntos es consistente de manera global con las observaciones de la escena capturada, al igual que lo son las poses estimadas de la cámara. Estos resultados también se aprecian de manera visual en las nubes de puntos, donde se observan estructuras que se asemejan a la escena real capturada. Es importante recordar que la triangulación es realizada sobre una estimación inicial de poses donde se desconoce la magnitud real de la traslación, por lo que es de esperar que la nube de puntos obtenida no represente la escena real. Así, entre las secuencias utilizadas, destaca el caso de la reconstrucción lograda en la secuencia 'HerzJesu', donde, teniendo un alto error de reproyección promedio inicial, el algoritmo de optimización converge a una reconstrucción precisa, como se observa en la Fig. 4.24, donde las magnitudes de las traslaciones de cámara quedan relativas a la magnitud de las traslación existente entre las primeras dos vistas. En resumen, la calidad de la reconstrucción aumenta considerablemente luego de la optimización global realizada, haciendo de la etapa de *Bundle Adjustment* un elemento esencial del algoritmo desarrollado.

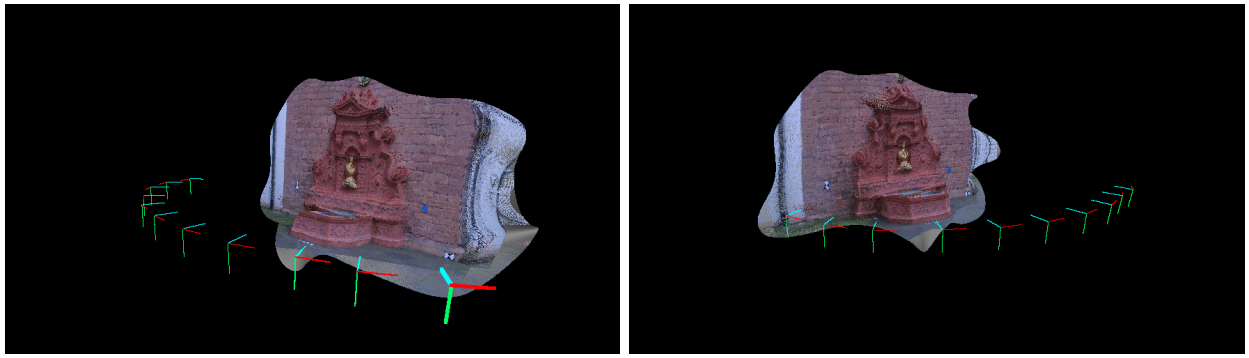
En base a los resultados obtenidos para el ajuste de superficie B-Spline, es posible afirmar que esta, en los casos analizados, logra representar la estructura de la escena de manera aproximada, ya que posee curvas suaves que no pueden representar muchas de las discontinuidades de la estructura real. En contraste con el método de registro por homografía global evaluado, que sólo permite relacionar el contenido en un plano, se considera que la aproximación de superficies B-Spline es de un nivel de detalle mucho mayor. Sin embargo, al ser ajustada sobre



(a) Reconstrucción parcial.



(b) Superficie ajustada para proyección.

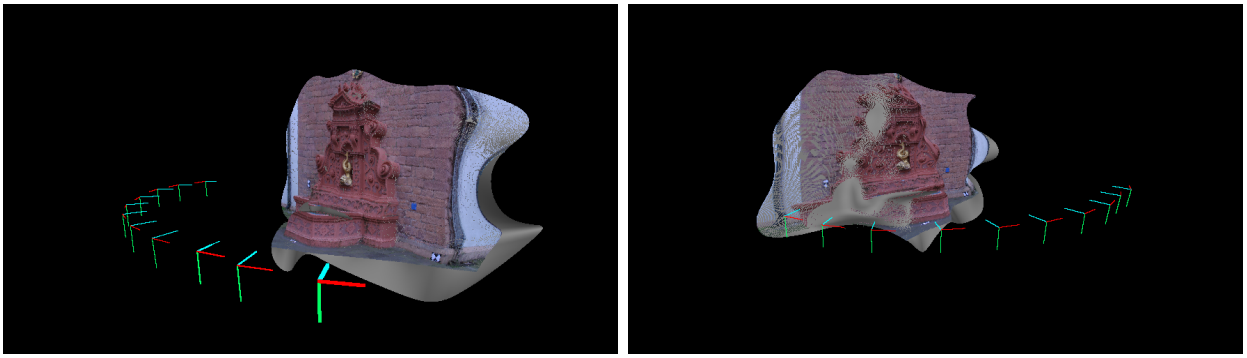


(c) Composición de imágenes sobre superficie

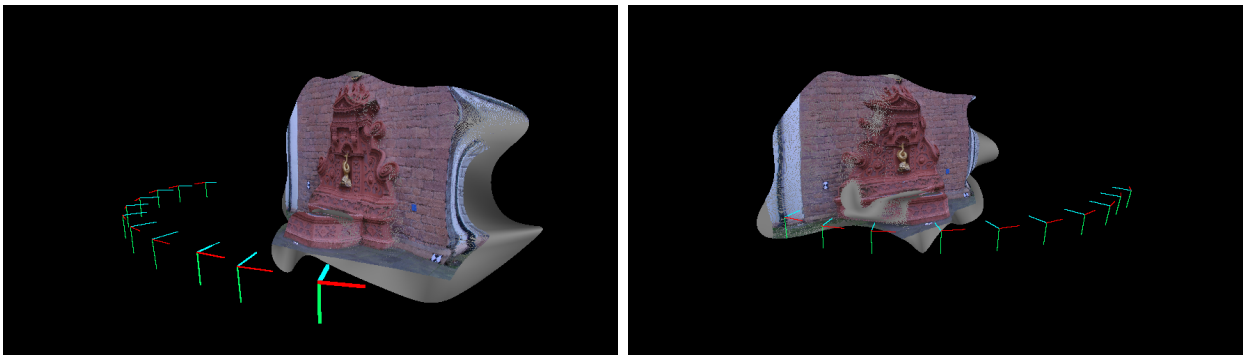
Figura 4.26: Resultado en dos vistas de la composición sobre modelo 3D, correspondiente a la secuencia 'Fountain'.

una reconstrucción parcial, existe una gran cantidad de zonas que no son representadas por la superficie, además de producir bordes deformados, como se ve claramente en la Fig. 4.25.

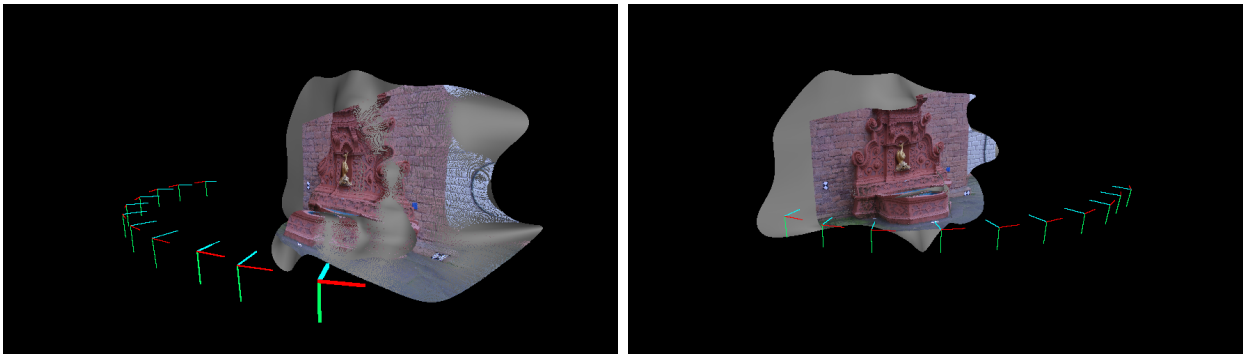
Posteriormente, proyectando las vistas como una imagen de rango desde las distintas poses de cámara, se observa que, en general, se logra ajustar el contenido de la imagen a la estructura aproximada por la superficie. En el caso de la secuencia 'Fountain', el contenido correspondiente a la fuente es proyectado correctamente, permitiendo obtener un modelo 3D coloreado que puede ser observado desde distintos puntos de vista. En la Fig. 4.27, se muestra claramente el contenido que aportan las distintas imágenes, el cual, siendo proyectado de manera independiente, queda incompleto debido a las zonas no visibles desde ese punto de vista. Esto demuestra que el algoritmo de stitching desarrollado, posee un gran potencial



(a) Proyección de primera imagen de la secuencia.



(b) Proyección de tercera imagen de la secuencia.

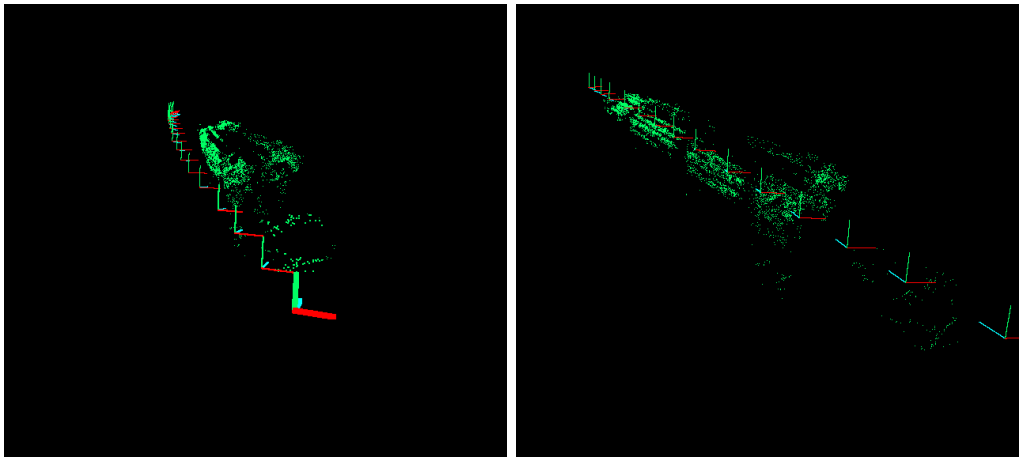


(c) Proyección de séptima imagen de la secuencia.

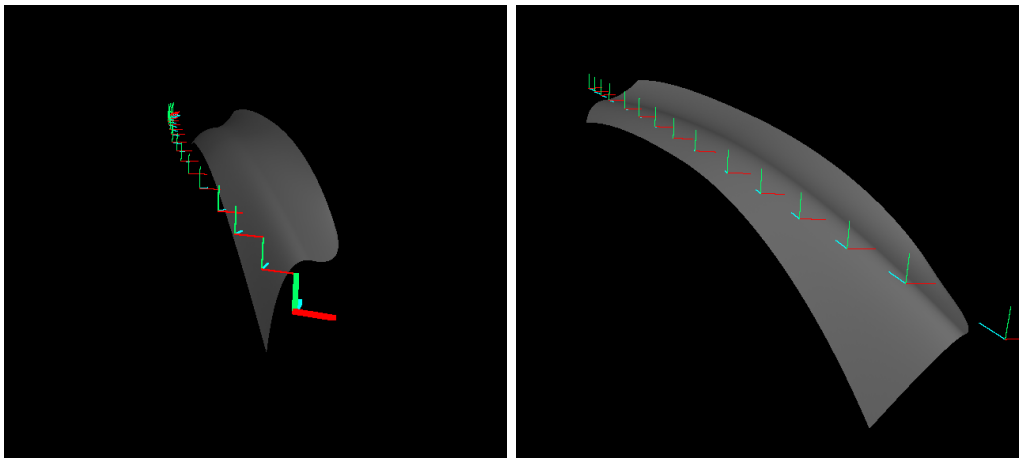
Figura 4.27: Resultado en dos vistas de proyecciones de imágenes de la secuencia sobre modelo 3D, correspondiente a la secuencia 'Fountain'.

para la fusión del contenido de imágenes en el caso de las secuencias con desplazamientos amplios de cámara.

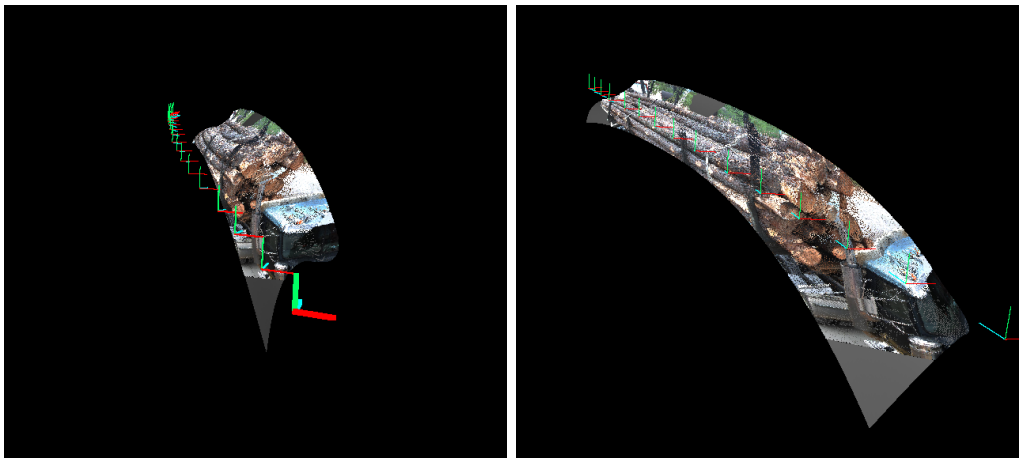
Antes de analizar los resultados obtenidos sobre la secuencia del camión, es importante mencionar que la complejidad de estas secuencias es superior a la de los casos anteriores, ya que los datos fueron extraídos de un producto en planta, con condiciones poco favorables, como lo es la tasa de captura de 1[fps], que dificulta establecer correspondencias densas al existir grandes cambios entre cada cuadro de la secuencia. Con esto en consideración, al observar los resultados de la Fig. 4.28, en primer lugar, se identifica que la nube de puntos reconstruida posee una baja densidad, haciendo difícil determinar que se trata de la visualización de un camión con cargamento de troncos. En segundo lugar, analizando la superficie



(a) Reconstrucción parcial.



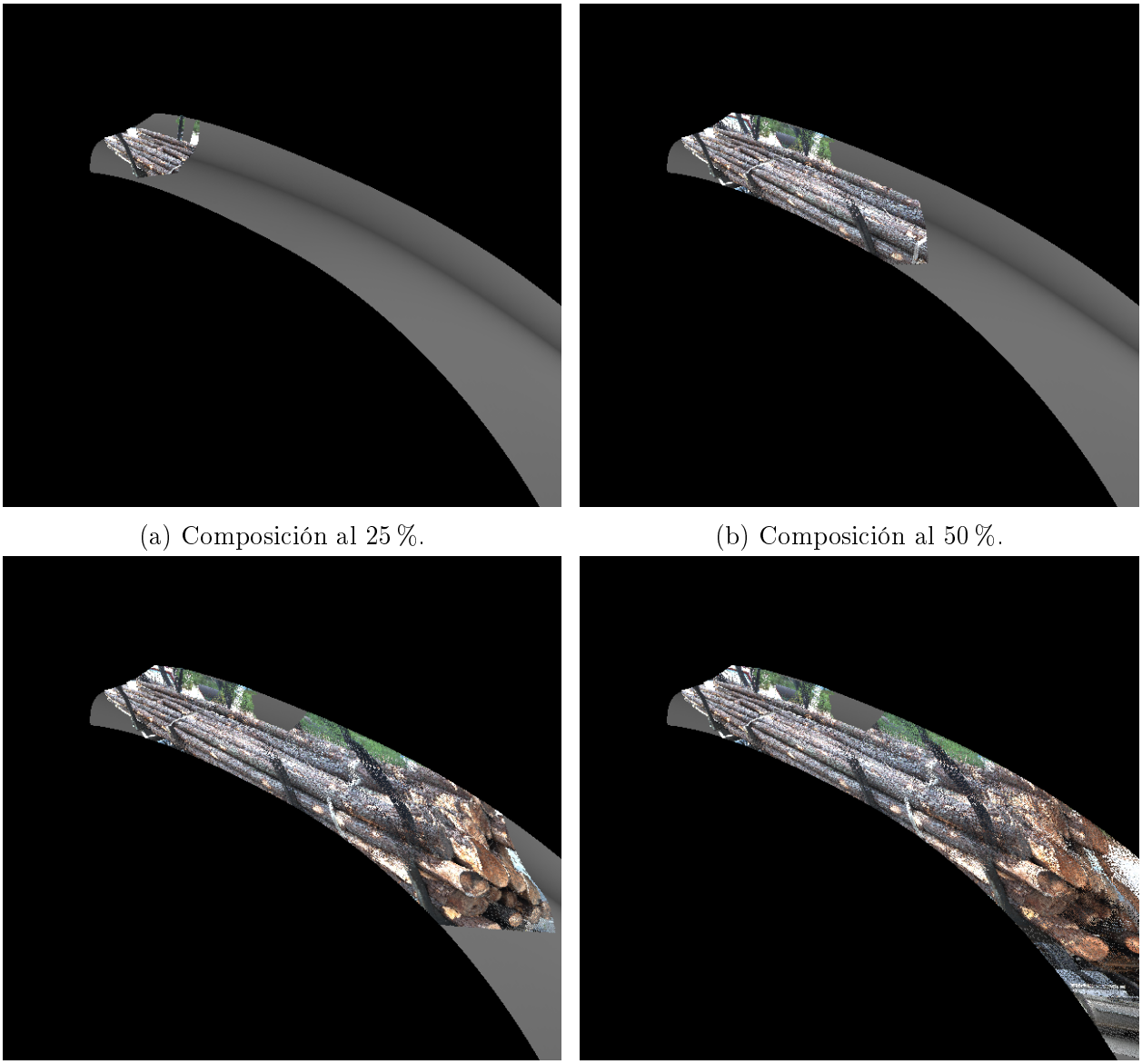
(b) Superficie ajustada para proyección.



(c) Composición de imágenes sobre superficie

Figura 4.28: Resultado en dos vistas de la composición sobre modelo 3D, correspondiente a la secuencia 'Truck2L'.

aproximada, se ve que esta representa un manto general sobre la nube de puntos, que se ajusta levemente a la zona superior y lateral superior del cargamento, pero no al sector de la cabina o a la parte lateral inferior de la carga. En este sentido, el modelo aproximado del



(a) Composición al 25 %. (b) Composición al 50 %.

(c) Composición al 75 %. (d) Composición al 100 %.

Figura 4.29: Resultados parciales de la composición al proyectar distintas imágenes de la secuencia 'Truck2L' sobre modelo 3D.

camión queda incompleto y no posee la calidad esperada para lograr una buena composición como en los casos anteriores. Sin embargo, con los resultados obtenidos se identifica que es posible lograr una mejora al modificar las condiciones de captura. Por ejemplo, aumentando la tasa de muestreo de la cámara, sería posible generar correspondencias más densas y así, mejorar la nube de puntos.

Por otro lado, al proyectar las imágenes del camión sucesivamente, como muestra la Fig. 4.29, se observa que se logra alinear el contenido correspondiente a las zonas de los troncos a las que se ajusta la superficie. En las zonas donde la superficie no se ajusta, como por ejemplo las barras verticales del cargamento, las caras frontales de los troncos y el fondo residual, las imágenes proyectadas quedan totalmente desalineadas, produciendo artefactos que distorsio-

nan la composición final. Esta situación da a conocer que para lograr un desarrollo acabado del algoritmo de stitching, que fusione las imágenes del *Logmeter*, se deben solucionar los tres problemas mencionados: la baja densidad de la nube de puntos, la generación de una superficie y la elección de zonas a proyectar, evitando aquellas que introduzcan artefactos.

Finalmente, a pesar de los errores de alineación, se identifica que el enfoque de stitching por reconstrucción de escena posee fuertes ventajas respecto del stitching por homografías, entre las cuales destaca el hecho de permitir la visualización de la composición desde cualquier punto de vista, lo que contrasta con la pérdida de información existente al proyectar una vista lejana sobre una vista de referencia mediante una homografía. Otra ventaja importante, corresponde a la obtención de una reconstrucción parcial de la escena, que posee el potencial de otorgar información valiosa para diversas aplicaciones que escapan al simple procesamiento de imágenes planas.

4.3.5. Discusión general del trabajo realizado

Con el objetivo de analizar de manera global el trabajo realizado, a continuación se presenta una discusión sobre la metodología abordada y las decisiones tomadas, considerando su impacto sobre los resultados finales obtenidos.

Primero, considerando la alta complejidad del problema planteado, se propuso un diseño de algoritmo de stitching basado en la extracción de características, abordando una etapa general de alineamiento y fusión que permite ser implementada de diversas maneras. El plan de trabajo definido involucra, en el inicio, una fase de exploración con un prototipo que realiza alineamiento mediante transformaciones homográficas, en el cual se consigue desarrollar las etapas de mejora de contraste, extracción de características, establecimiento y validación de calces, además de la estimación de homografías y posterior proyección y composición. Basado en el estudio bibliográfico realizado, se escoge AKAZE como extractor de características, lo cual, sumado a la mejora de contraste, permite obtener correspondencias entre puntos de interés que abarcan completamente cada imagen de las secuencias tratadas. Si bien existen descriptores que se comportan mejor en casos más complejos, como lo es A-SIFT [36], se considera que para efectos de implementar la propuesta de algoritmo de stitching, el extractor AKAZE es suficiente. De la misma forma, se encuentra en CLAHE, un método de mejora de contraste efectivo, el cual, sumado a otras técnicas de pre-procesamiento, podrían incrementar la cantidad y densidad de las correspondencias obtenidas.

En segundo lugar, la decisión de implementar un prototipo del algoritmo de stitching por homografías permite no sólo conocer las limitaciones del mismo sobre las secuencias de imágenes abordadas, si no que también da la base que permite implementar el algoritmo de stitching por reconstrucción de escenas. En el prototipo, se demuestra que las homografías son válidas sólo para una zona acotada de la imagen, correspondiente a un plano, por lo que al usar secuencias con estructuras complejas y desplazamientos de cámara, las distorsiones en la composición final son demasiado grandes. Por otro lado, en la propuesta de algoritmo de stitching por reconstrucción de escenas, se incorporan técnicas pertenecientes al estado del arte en el flujo de reconstrucción. Si bien, los resultados obtenidos indican que se requiere mayor desarrollo, se logra validar conceptualmente la propuesta de mezclar las imágenes sobre

un modelo 3D. En el caso del ajuste de superficie, basado en el trabajo de [66], se decidió usar superficies B-Spline. En el trabajo del autor, la superficie es utilizada con éxito para proyectar la imagen de rango de una sola vista y así obtener la perspectiva frontal de un automóvil. No obstante, su aplicación en el presente trabajo demuestra que existe un gran potencial para su uso en algoritmos de stitching. Posibles mejoras en este aspecto podrían incluir el ajuste y mezclado de distintas superficies B-Spline, de manera de que cada superficie represente por separado zonas muy diferentes, como por ejemplo, la cabina y la carga del camión en las secuencias del *Logmeter*, usando un esquema análogo al registro de homografías duales de [44].

Finalmente, el trabajo realizado para Woodtech S.A., además de generar una propuesta para el algoritmo de stitching sobre secuencias de imágenes con amplios movimientos de cámara, se introducen las etapas necesarias para adaptarlo a las secuencias de imágenes capturadas en el *Logmeter*. En este sentido, se logran resultados similares al trabajo de [66], respecto de la determinación de poses de cámara relativas al movimiento de un vehículo y la reconstrucción parcial de la escena, optimizada con *Bundle Adjustment*.

Capítulo 5

Conclusión

En este trabajo, se desarrolla una propuesta de un algoritmo de stitching para sintetizar la información desplegada en imágenes de una secuencia que cuenta con amplios desplazamientos de cámara durante su captura. Adicionalmente, se logra plantear una forma de adaptar el diseño al caso de secuencias obtenidas de una cámara estática que captura un vehículo en movimiento, considerando el problema dual correspondiente al movimiento relativo de la cámara respecto a un vehículo estático, permitiendo probar el algoritmo desarrollado sobre imágenes capturadas por el producto *Logmeter* de la empresa Woodtech S.A..

Luego, considerando los objetivos iniciales, se concluye que el trabajo realizado cumple los alcances del proyecto, generando una propuesta que se fundamenta conceptualmente en técnicas del estado del arte para resolver el problema abordado. En este contexto, se identifica la necesidad de realizar etapas adicionales de desarrollo para ampliar las capacidades de la implementación existente y, así, generar una versión que se pueda incorporar en un producto comercial. En este sentido, al implementar el algoritmo mediante bloques de *software* independientes, el resultado de este trabajo es fácilmente extensible a otras aplicaciones relacionadas, como lo son, por ejemplo, el stitching sobre secuencias con distintas características, o la reconstrucción tridimensional de camiones y su carga usando una o más cámaras en productos existentes o en el diseño de nuevos productos.

La metodología de la propuesta final, incorpora técnicas de reconstrucción de escenas, como la triangulación y optimización simultánea de las poses de la cámara y de la nube de puntos reconstruida, mediante *Bundle Adjustment*. En este contexto, las pruebas realizadas muestran una reducción del error de reproyección promedio, llegando a valores bajo los 0,01$[px]$ en todas las secuencias, lo cual se traduce en que la reconstrucción parcial es consistente de manera global con las observaciones de la escena capturada, generando nubes de puntos que visualmente se asemejan a la escena real capturada. Por otro lado, mediante la evaluación visual de los resultados, se valida conceptualmente el diseño propuesto, identificando una mejor capacidad de alineamiento para las escenas complejas en el tipo de secuencias tratadas, respecto del prototipo mediante alineamiento por transformaciones homográficas. Adicionalmente, se concluye que el uso de superficies B-spline, si bien es una aproximación de la escena, contribuye a definir una superficie de proyección que permite combinar las imágenes en ella.

Finalmente, se concluye que el desarrollo realizado entrega una base para las líneas de trabajo del equipo de desarrollo de la empresa Woodtech S.A., correspondientes a la fusión de imágenes para la inspección de carga y la reconstrucción tridimensional del camión mediante captura de imágenes desde cámaras. Con este trabajo, se da al equipo la posibilidad de mejorar el algoritmo propuesto, o basarse en los principios utilizados para la reconstrucción de escenas, buscando obtener una reconstrucción densa del camión, dando pie a complementar o incluso reemplazar la estructura obtenida mediante medición láser.

5.1. Trabajo futuro

Como trabajo futuro, se consideran dos líneas de desarrollo para dar continuidad al trabajo realizado. En primer lugar, una línea asociada a buscar mejorar la visualización de la composición final obtenida como el resultado de la proyección de imágenes sobre una superficie y, en segundo lugar, una línea de desarrollo orientada a lograr una reconstrucción densa de la escena, con una calidad tal, que no requiera de proyectar imágenes sobre una superficie, si no que la reconstrucción misma sea la que sintetice la información de las vistas.

En la primera línea, se podría investigar otras formas de generar una superficie que represente de manera fiel la nube de puntos correspondiente a la estructura reconstruida. Por otro lado, se podría buscar que la composición sobre dicha superficie, considere la selección adecuada de las zonas a proyectar de cada imagen, buscando obtener un resultado que reduzcan la cantidad de artefactos generados por zonas mal alineadas.

Por otro lado, en la segunda línea, se plantea como posibilidad, el uso de algoritmos de *Multi-view Stereo*, basados en el registro de parches de superficies 3D coloreadas [53], logrando una reconstrucción densa que no requiere del ajuste de una superficie y que lleva directamente la información de las imágenes al modelo reconstruido. Otra posibilidad está en aprovechar las poses de cámara registradas luego de la optimización, para realizar una nueva búsqueda de correspondencias, aprovechando esta nueva información para utilizar un descriptor que permita relacionar zonas más densas.

Por último, dado que el desarrollo realizado logra una reconstrucción parcial de la escena en una escala indeterminada, en un trabajo futuro, se podría incorporar información externa para determinar la escala apropiada y obtener una reconstrucción métrica de las escenas. Esta línea de investigación es particularmente relevante para la empresa Woodtech S.A., ya que al incorporar, por ejemplo, el desplazamiento del camión en movimiento y lograr una reconstrucción métrica del cargamento, se podrían realizar mediciones con esta información, privilegiando el uso de las cámaras respecto a los sensores láser para reducir costos en una versión simplificada del producto.

Bibliografía

[1] Red to green website. Red To Green S.A. [Online]. Available: http://www.redtogreen.com/

[2] Logmeter website. Woodtech S.A. [Online]. Available: http://www.woodtechms.com/

[3] M. Nylinder, T. Kubénka, and M. Hultnäs, "Roundwood measurement of truck loads by laser scanning," *Field study at Arauco pulp mill Nueva Aldea*, pp. 1–9, 2008.

[4] Imágenes de ejemplo con cargamento de madera capturadas por producto *Logmeter*. Woodtech S.A.

[5] R. Szeliski, *Computer Vision: Algorithms and Applications*, 1st ed. New York, NY, USA: Springer-Verlag New York, Inc., 2010.

[6] M. Brown and D. G. Lowe, "Recognising panoramas," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, Oct 2003, pp. 1218–1225 vol.2.

[7] N. Gracias, M. Mahoor, S. Negahdaripour, and A. Gleason, "Fast image blending using watersheds and graph cuts," *Image and Vision Computing*, vol. 27, no. 5, pp. 597–607, 2009.

[8] F. Zhang and F. Liu, "Parallax-tolerant image stitching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3262–3269.

[9] L. Rosado, J. Gonçalves, J. Costa, D. Ribeiro, and F. Soares, "Supervised learning for out-of-stock detection in panoramas of retail shelves," in *2016 IEEE International Conference on Imaging Systems and Techniques (IST)*, Oct 2016, pp. 406–411.

[10] X. Shi, X. Huang, and D. Zhang, "Fast stitch algorithm on aerial images," in *Mechatronic Sciences, Electric Engineering and Computer (MEC), Proceedings 2013 International Conference on*, Dec 2013, pp. 2213–2217.

[11] G. Zhang, Y. He, W. Chen, J. Jia, and H. Bao, "Multi-Viewpoint Panorama Construction with Wide-Baseline Images," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3099–3111, 2016.

[12] E. Adel, M. Elmogy, and H. Elbakry, "Image Stitching based on Feature Extraction Techniques: A Survey," *International Journal of Computer Application*, vol. 99, no. 6,

pp. 1–8, 2014.

[13] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. New York, NY, USA: Cambridge University Press, 2003.

[14] G. B. Gene Cooper. Camera motions for stitching. Make: magazine. [Online]. Available: http://i0.wp.com/makezine.com/wp-content/uploads/2014/03/rotation_translation.jpg

[15] J. Lobo and J. Dias, "Relative pose calibration between visual and inertial sensors," *The International Journal of Robotics Research*, vol. 26, no. 6, pp. 561–575, 2007.

[16] D. Nister, O. Naroditsky, and J. Bergen, "Visual odometry," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1, June 2004, pp. I–652–I–659 Vol.1.

[17] Clahe results on retinal image. Developers Club Blog. [Online]. Available: https://hsto.org/files/4e8/a3d/057/4e8a3d05785b4a6d9467264e4f3c0c28.jpg

[18] B. S. Min, D. K. Lim, S. J. Kim, and J. H. Lee, "A novel method of determining parameters of CLAHE based on image entropy," *International Journal of Software Engineering and its Applications*, vol. 7, no. 5, pp. 113–120, 2013.

[19] S. Beucher and F. Meyer, "The morphological approach to segmentation: the watershed transformation," *Optical Engineering*, vol. 34, pp. 433–433, 1992.

[20] M. Piccardi, "Background subtraction techniques: a review," in *Systems, man and cybernetics, 2004 IEEE international conference on*, vol. 4. IEEE, 2004, pp. 3099–3104.

[21] I. Kartika and S. S. Mohamed, "Frame differencing with post-processing techniques for moving object detection in outdoor environment," in *Signal Processing and its Applications (CSPA), 2011 IEEE 7th International Colloquium on*. IEEE, 2011, pp. 172–176.

[22] Y. Zhang, X. Wang, and B. Qu, "Three-frame difference algorithm research based on mathematical morphology," *Procedia Engineering*, vol. 29, pp. 2705–2709, 2012.

[23] A. Kumar, J. M. Hart, and N. Ahuja, "Motion-based background subtraction and panoramic mosaicing for freight train analysis," in *2013 IEEE International Conference on Image Processing*, Sept 2013, pp. 4564–4568.

[24] J. Bernal, F. Vilarino, and J. Sánchez, "Feature detectors and feature descriptors: where we are now," *Computer Vision Center and Computer Science Department UAB Campus UAB, Edi_ ci O*, vol. 8193, 2010.

[25] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International journal of computer vision*, vol. 74, no. 1, pp. 59–73, 2007.

[26] J. Shi and C. Tomasi, "Good features to track," in *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on.*

IEEE, 1994, pp. 593–600.

[27] B. Triggs, "Detecting keypoints with stable position, orientation, and scale under illumination changes," in *European conference on computer vision*. Springer, 2004, pp. 100–113.

[28] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *European conference on computer vision*. Springer, 2006, pp. 430–443.

[29] M. Agrawal, K. Konolige, and M. R. Blas, "Censure: Center surround extremas for realtime feature detection and matching," in *European Conference on Computer Vision*. Springer, 2008, pp. 102–115.

[30] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[31] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*. Springer, 2006, pp. 404–417.

[32] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *2011 International Conference on Computer Vision*, Nov 2011, pp. 2564–2571.

[33] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "Kaze features," in *European Conference on Computer Vision*. Springer, 2012, pp. 214–227.

[34] S. Grewenig, J. Weickert, and A. Bruhn, *From Box Filtering to Fast Explicit Diffusion*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 533–542. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-15986-2_54

[35] P. F. Alcantarilla and T. Solutions, "Fast explicit diffusion for accelerated features in nonlinear scale spaces," *IEEE Trans. Patt. Anal. Mach. Intell*, vol. 34, no. 7, pp. 1281–1298, 2011.

[36] J.-M. Morel and G. Yu, "Asift: A new framework for fully affine invariant image comparison," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 438–469, 2009.

[37] E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 815–830, May 2010.

[38] P.-E. Forssén and D. G. Lowe, "Shape descriptors for maximally stable extremal regions," in *2007 IEEE 11th International Conference on Computer Vision*. IEEE, 2007, pp. 1–8.

[39] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *International journal of computer vision*, vol. 65, no. 1-2, pp. 43–72, 2005.

[40] E. Adel, M. Elmogy, and H. Elbakry, "Real time image mosaicing system based on feature

extraction techniques," in *2014 9th International Conference on Computer Engineering Systems (ICCES)*, Dec 2014, pp. 339–345.

[41] M. Muja and D. G. Lowe, "Fast matching of binary features," in *Computer and Robot Vision (CRV)*, 2012, pp. 404–410.

[42] Homography. OpenMVG. [Online]. Available: http://imagine.enpc.fr/~moulonp/openMVG/Homography_geometry.png

[43] S. Šegvić, M. Ševrović, G. Kos, V. Stanisavljević, and I. Dadić, "Preliminary experiments in multi-view video stitching," in *2011 Proceedings of the 34th International Convention MIPRO*, May 2011, pp. 892–896.

[44] J. Gao, S. J. Kim, and M. S. Brown, "Constructing image panoramas using dual-homography warping," in *CVPR 2011*, June 2011, pp. 49–56.

[45] G. Zhang, Y. He, W. Chen, J. Jia, and H. Bao, "Multi-viewpoint panorama construction with wide-baseline images," *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3099–3111, July 2016.

[46] Essential matrix. OpenMVG. [Online]. Available: http://imagine.enpc.fr/~moulonp/openMVG/Essential_geometry.png

[47] P. F. McLauchlan and A. Jaenicke, "Image mosaicing using sequential bundle adjustment," *Image and Vision computing*, vol. 20, no. 9, pp. 751–759, 2002.

[48] Z. Kukelova, M. Bujnak, and T. Pajdla, "Closed-form solutions to minimal absolute pose problems with known vertical direction," in *Asian Conference on Computer Vision*. Springer, 2010, pp. 216–229.

[49] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[50] E. Dubrofsky, "Homography Estimation," *Optical Engineering*, vol. 15, no. March, p. 977, 2009. [Online]. Available: https://www.cs.ubc.ca/grads/resources/thesis/May09/Dubrofsky_Elan.pdf

[51] J. Li, P. Duan, and J. Wang, "Binocular stereo vision calibration experiment based on essential matrix," in *2015 IEEE International Conference on Computer and Communications (ICCC)*, Oct 2015, pp. 250–254.

[52] P. Li, D. Farin, R. K. Gunnewiek, and P. H. N. de With, "On creating depth maps from monoscopic video using structure from motion," in *Proc. IEEE Workshop on Content Generation and Coding for 3D-television*, 2006.

[53] Y. Furukawa, C. Hernández *et al.*, "Multi-view stereo: A tutorial," *Foundations and Trends ® in Computer Graphics and Vision*, vol. 9, no. 1-2, pp. 1–148, 2015.

[54] E. Marchand, H. Uchiyama, and F. Spindler, "Pose estimation for augmented reality: A hands-on survey," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 12, pp. 2633–2651, Dec 2016.

[55] D. Nister, "An Efficient Solution to the Five-Point Relative Pose Problem," vol. 26, no. 6, pp. 756–770, 2004.

[56] M. Sainz, N. Bagherzadeh, and A. Susin, "Recovering 3D metric structure and motion from multiple uncalibrated cameras," *Proceedings - International Conference on Information Technology: Coding and Computing, ITCC 2002*, pp. 268–273, 2002.

[57] H. Bazargani, "Camera Calibration and Pose Estimation from Planes," no. December, 2015.

[58] M. Bujnak, Z. Kukelova, and T. Pajdla, "New efficient solution to the absolute pose problem for camera with unknown focal length and radial distortion," in *Asian Conference on Computer Vision*. Springer, 2010, pp. 11–24.

[59] R. I. Hartley and P. Sturm, "Triangulation," *Computer vision and image understanding*, vol. 68, no. 2, pp. 146–157, 1997.

[60] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, 2006, pp. 519–528. [Online]. Available: http://vision.middlebury.edu/mview/data/

[61] C. Strecha, W. Von Hansen, L. Van Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. Ieee, 2008, pp. 1–8. [Online]. Available: http://cvlabwww.epfl.ch/data/multiview/denseMVS.html

[62] K. Kanatani and Y. Sugaya, "Bundle Adjustment for 3-D Reconstruction: Implementation and Evaluation," *Memoirs of the Faculty of ...*, vol. 45, no. January, pp. 1–9, 2011. [Online]. Available: http://www.suri.it.okayama-u.ac.jp/{~}kanatani/papers/okabundle.pdf

[63] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—a modern synthesis," in *International workshop on vision algorithms*. Springer, 1999, pp. 298–372.

[64] M. H. M. Patel, A. P. P. J. Patel, and A. P. M. S. G. Patel, "Comprehensive study and review of image mosaicing methods," in *International Journal of Engineering Research and Technology*, vol. 1. ESRSA Publications, 2012.

[65] H. I. Koo, B. S. Kim, and N. I. Cho, "A new method to find an optimal warping function in image stitching," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2009, pp. 1289–1292.

[66] A. Auclair, L. Cohen, and N. Vincent, "A robust approach for 3D cars reconstruction," *Proceedings of the 15th Scandinavian conference on Image analysis*, pp. 183–192, 2007. [Online]. Available: http://portal.acm.org/citation.cfm?id=1768615.1768639

[67] T. Weibel, C. Daul, D. Wolf, and R. Rösch, "Contrast-enhancing seam detection and blending using graph cuts," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Nov 2012, pp. 2732–2735.

[68] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in nd images," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 1. IEEE, 2001, pp. 105–112.

[69] Open source computer vision library. OpenCV. [Online]. Available: http://docs.opencv.org/3.1.0/d1/dfb/intro.html

[70] Point cloud library. PCL. [Online]. Available: http://pointclouds.org/

[71] Fitting trimmed b-splines to unordered point clouds. PCL Documentation. [Online]. Available: http://pointclouds.org/documentation/tutorials/bspline_fitting.php

[72] M. A. Lourakis and A. Argyros, "SBA: A Software Package for Generic Sparse Bundle Adjustment," *ACM Trans. Math. Software*, vol. 36, no. 1, pp. 1–30, 2009.

[73] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, "MeshLab: an Open-Source Mesh Processing Tool," in *Eurographics Italian Chapter Conference*, V. Scarano, R. D. Chiara, and U. Erra, Eds. The Eurographics Association, 2008.

[74] Nurbs surface with control points. 3ds-Max tutorials. [Online]. Available: http://www.3dmax-tutorials.com/graphics/il_nurbs_cvsurf.jpg

[75] P. Moulon. (2012) Image of çhateau de sceaux", sceaux castle. france. [Online]. Available: https://github.com/openMVG/ImageDataset_SceauxCastle/tree/master/images

Capítulo 6

Anexos

6.1. Bases de datos

Se resumen en la Tabla 6.1, los aspectos técnicos de cada secuencia, como la resolución de las imágenes utilizadas y los parámetros intrínsecos de las cámaras, además de la transformación realizada respecto de las imágenes de la base de datos original de donde se obtuvieron. Las secuencias 'Fountain', 'HerzJesu' y 'SceauxCastle' fueron utilizadas a media resolución debido a la limitación en capacidad de procesamiento de la implementación desarrollada. Por otro lado, la matriz de parámetros intrínsecos de las secuencias del *Logmeter*, es construida utilizando la información del sensor de la cámara, sin considerar un modelo de distorsión del lente. Luego, desde la Fig. 6.1 hasta la Fig. 6.3, se exponen las imágenes de las secuencias escogidas.

Secuencia	**Largo**	**Resolución** (*Ancho* × *Alto*)	**Matriz de parámetros intrínsecos (K)**	**Transformación**
'Temple'	12	480 × 640	$\begin{bmatrix} 1520{,}4 & 0{,}0 & 246{,}87 \\ 0{,}0 & 1525{,}9 & 302{,}32 \\ 0{,}0 & 0{,}0 & 1{,}0 \end{bmatrix}$	Rotada en 90°.
'Fountain'	11	1536 × 1024	$\begin{bmatrix} 1379{,}74 & 0{,}0 & 774{,}66 \\ 0{,}0 & 1382{,}08 & 503{,}41 \\ 0{,}0 & 0{,}0 & 1{,}0 \end{bmatrix}$	Media resolución.
'HerzJesu'	14	1536 × 1024	$\begin{bmatrix} 1379{,}74 & 0{,}0 & 774{,}66 \\ 0{,}0 & 1382{,}08 & 503{,}41 \\ 0{,}0 & 0{,}0 & 1{,}0 \end{bmatrix}$	Media resolución.
'SceauxCastle'	11	1416 × 1064	$\begin{bmatrix} 1452{,}94 & 0{,}0 & 707{,}0 \\ 0{,}0 & 1452{,}94 & 532{,}0 \\ 0{,}0 & 0{,}0 & 1{,}0 \end{bmatrix}$	Media resolución.
'Truck1R'	21	1388 × 1038	$\begin{bmatrix} 1520{,}4 & 0{,}0 & 246{,}87 \\ 0{,}0 & 1525{,}9 & 302{,}32 \\ 0{,}0 & 0{,}0 & 1{,}0 \end{bmatrix}$	-
'Truck1L'	18	1388 × 1038	$\begin{bmatrix} 1520{,}4 & 0{,}0 & 246{,}87 \\ 0{,}0 & 1525{,}9 & 302{,}32 \\ 0{,}0 & 0{,}0 & 1{,}0 \end{bmatrix}$	-
'Truck2R'	14	1388 × 1038	$\begin{bmatrix} 1520{,}4 & 0{,}0 & 246{,}87 \\ 0{,}0 & 1525{,}9 & 302{,}32 \\ 0{,}0 & 0{,}0 & 1{,}0 \end{bmatrix}$	-
'Truck2L'	19	1388 × 1038	$\begin{bmatrix} 1520{,}4 & 0{,}0 & 246{,}87 \\ 0{,}0 & 1525{,}9 & 302{,}32 \\ 0{,}0 & 0{,}0 & 1{,}0 \end{bmatrix}$	-
'PaisajeNavarino'	2	1632 × 1224	-	-

Tabla 6.1: Resumen de bases de datos escogidas, con detalle de la resolución de las imágenes utilizadas y los parámetros intrínsecos de las cámaras, además de la transformación realizada respecto de las imágenes de la base de datos original de donde se obtuvieron las secuencias.

Figura 6.1: Secuencia de imágenes ‘Temple’.

Figura 6.2: Secuencia de imágenes 'Fountain'.

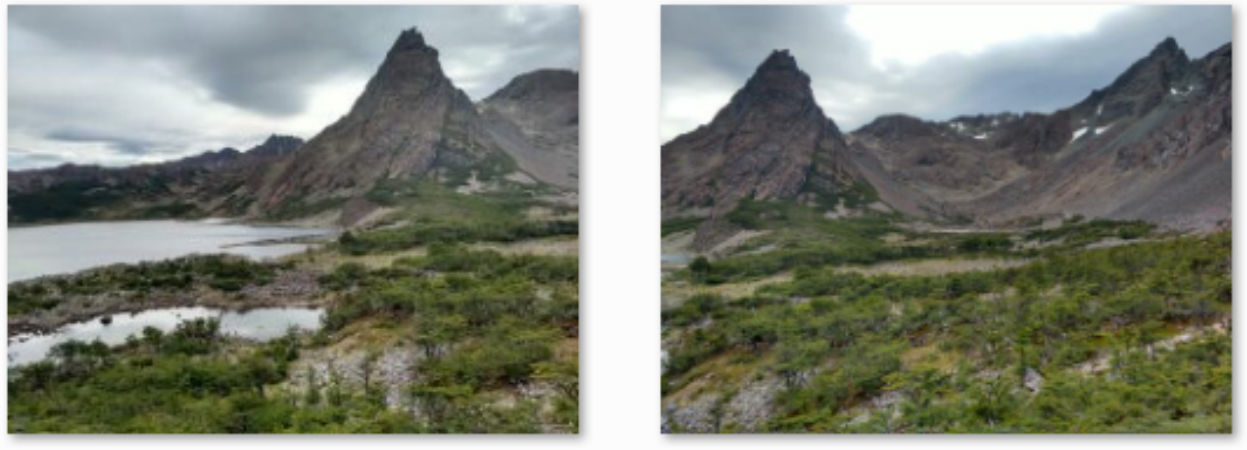


Figura 6.3: Secuencia de imágenes 'PaisajeNavarino'.

Figura 6.4: Secuencia de imágenes ‘HerzJesu’.

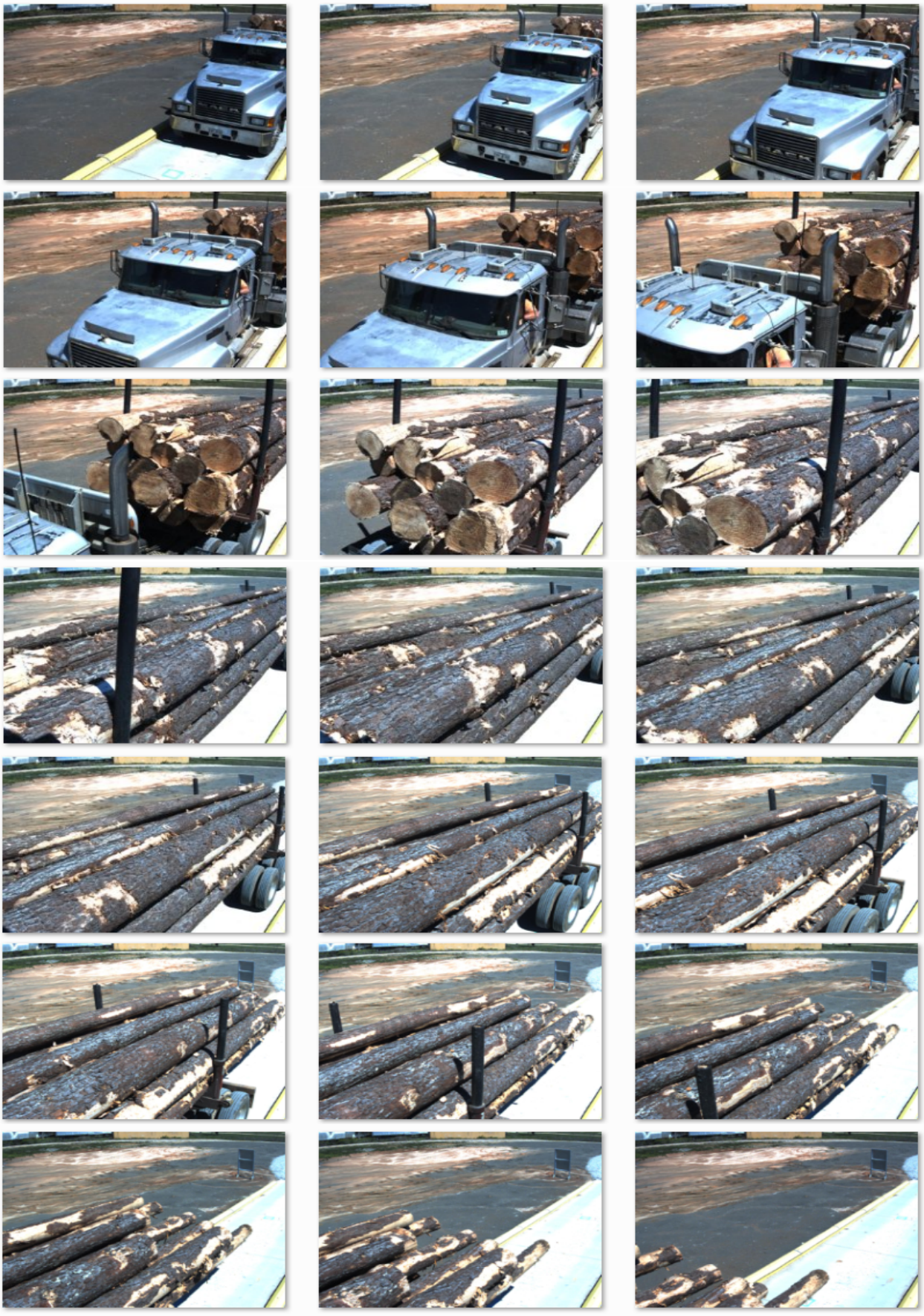


Figura 6.5: Secuencia de imágenes 'Truck1R'.

Figura 6.6: Secuencia de imágenes 'Truck1L'.

Figura 6.7: Secuencia de imágenes 'Truck2R'.

Figura 6.8: Secuencia de imágenes 'Truck2L'.

6.2. Efecto de mejora de contraste CLAHE

La siguiente tabla, resume la evaluación del efecto que logra la mejora de contraste CLAHE, en las distintas secuencias.

	Detección de puntos de interés	**Total Correspondencias**		
Secuencia	**Aumento Promedio [%]**	**Sin CLAHE**	**Con CLAHE**	**Aumento [%]**
'Temple'	43,96	4583	6358	38,73
'Fountain'	818,24	4576	51043	1015,45
'HerzJesu'	345,20	8621	24339	182,32
'SceauxCastle'	271,21	8694	35488	308,19
'Truck1R'	65,34	10060	12182	21,09
'Truck1L'	26,35	7942	7488	-5,72
'Truck2R'	79,62	5514	7381	33,86
'Truck2L'	54,32	10167	11967	17,70

Tabla 6.2: Efecto de mejora de contraste CLAHE sobre cantidad de puntos de interés detectados y total de correspondencias validadas por secuencia

6.3. Resultados adicionales del prototipo de algoritmo de stitching por homografías

A continuación, se muestran los resultados adicionales de la prueba realizada en el prototipo y los resultados obtenidos en la composición por *blending*, para las secuencias 'Temple', 'Truck1R' y 'Truck2L'.

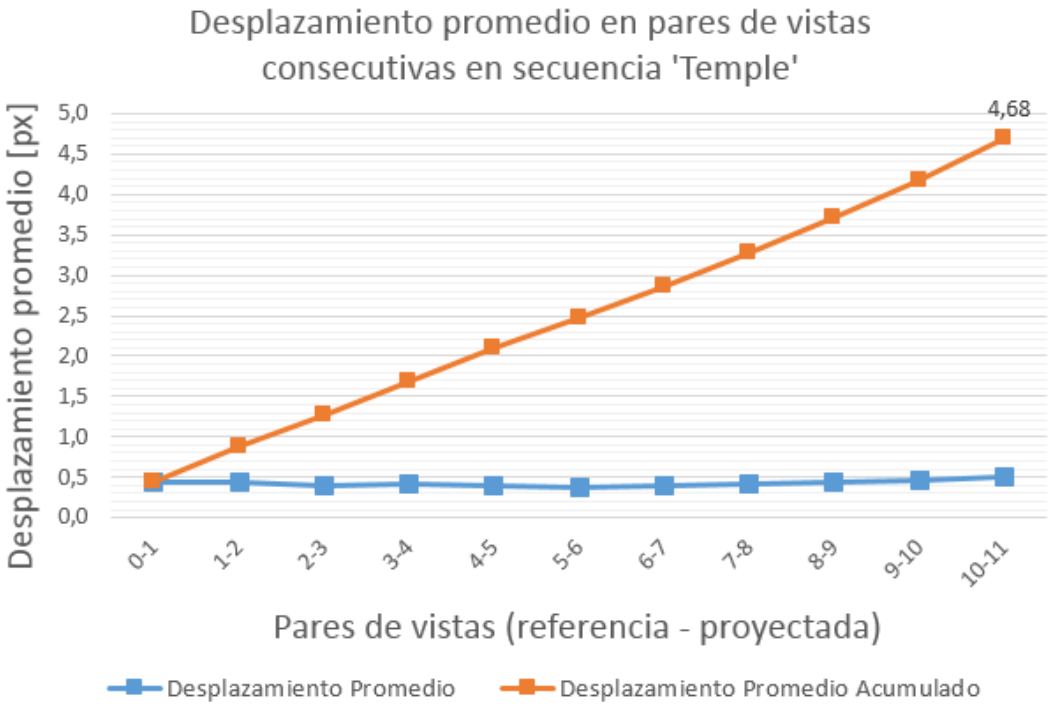


Figura 6.9: Desplazamiento promedio en pares de vistas consecutivas en secuencia 'Temple'. El desplazamiento acumulado para llevar la última vista hacia la primera es de 4,68[px]

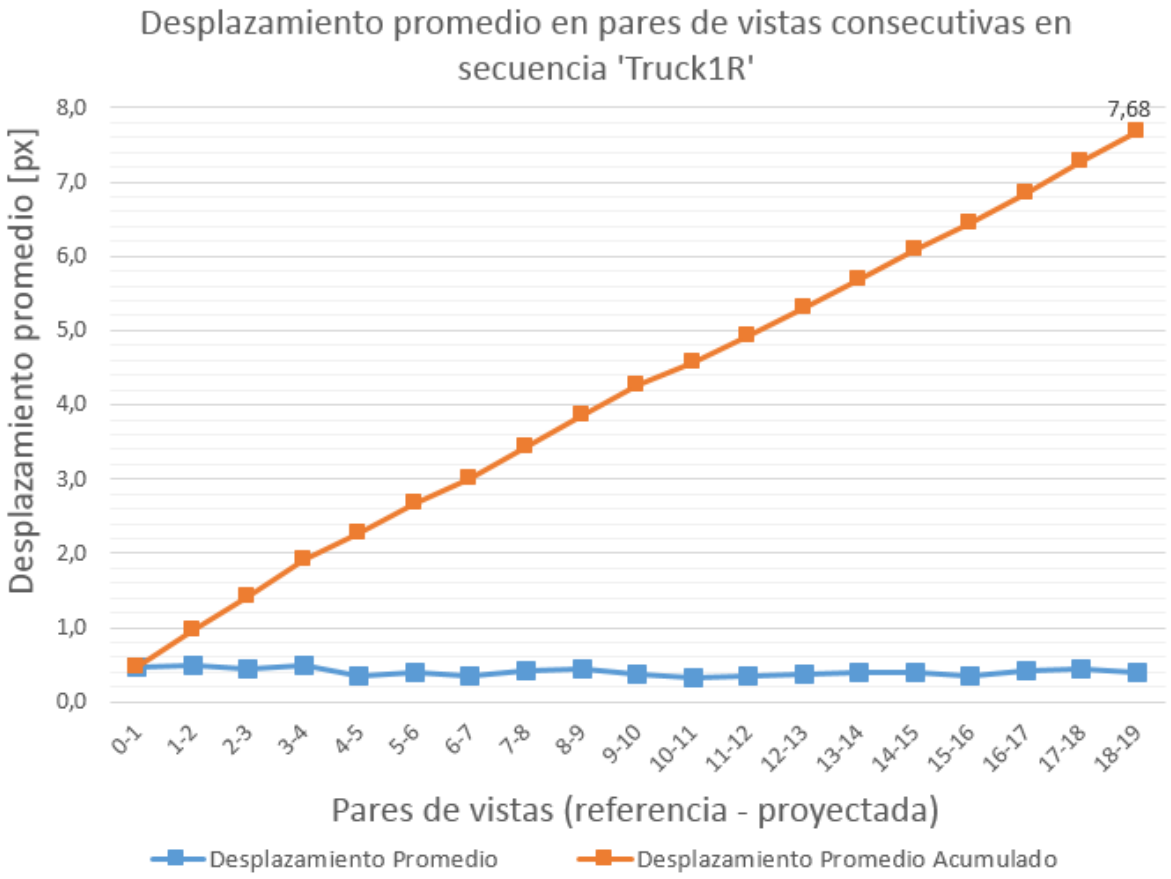
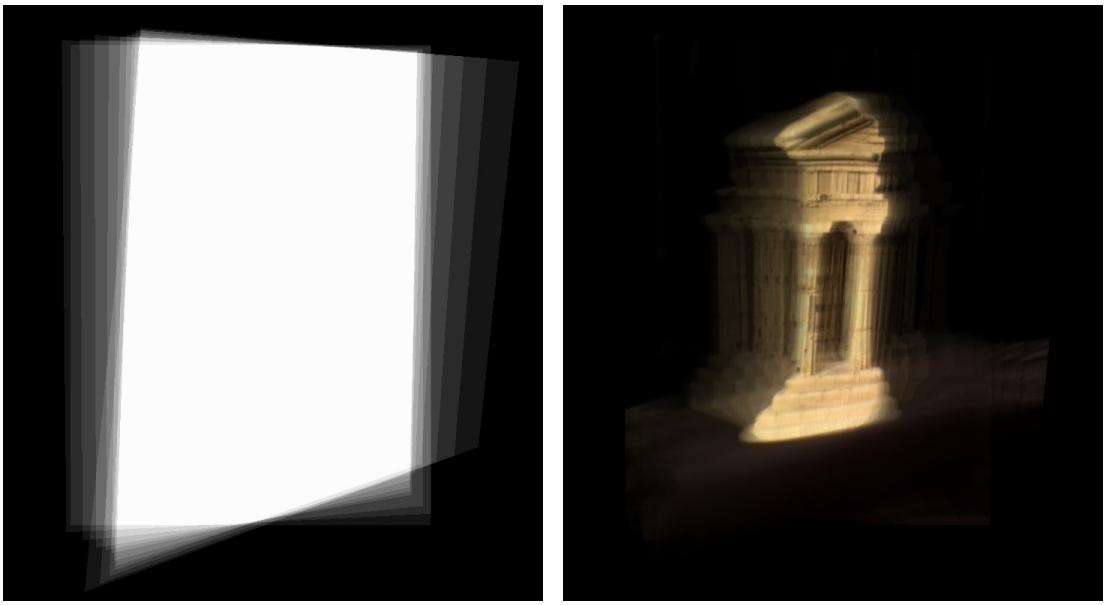
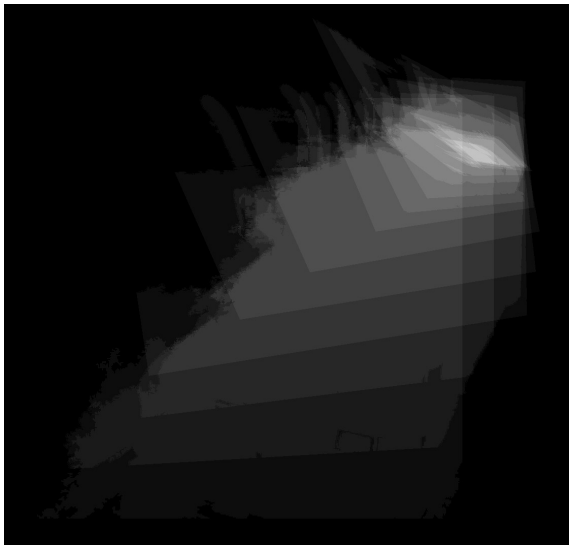


Figura 6.10: Desplazamiento promedio en pares de vistas consecutivas en secuencia 'Truck1R'. El desplazamiento acumulado para llevar la última vista hacia la primera es de 7,68[px]

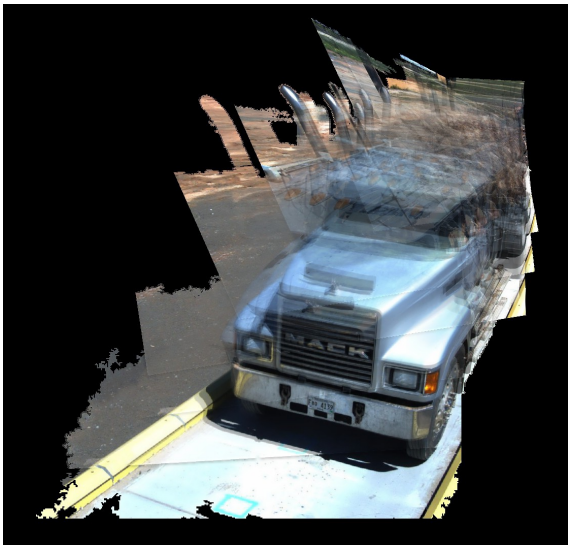


(a) Proyección de máscaras. (b) Composición por *blending* lineal.

Figura 6.11: Resultado de la composición de la secuencia 'Temple', mediante registro por homografías globales hacia el plano de referencia de la primera imagen.

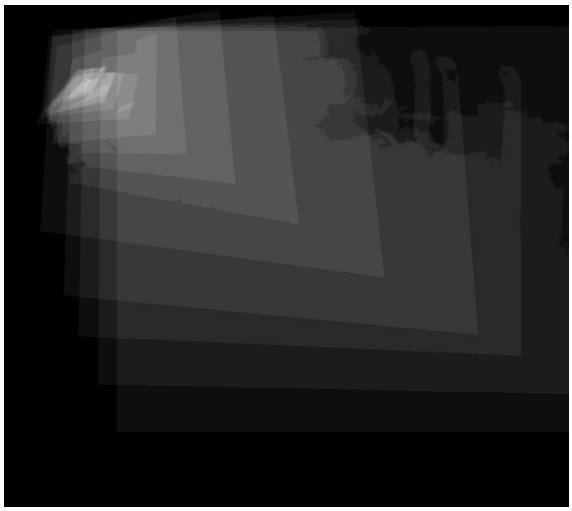


(a) Proyección de máscaras.



(b) Composición por *blending* lineal.

Figura 6.12: Resultado de la composición de la secuencia 'Truck1R', mediante registro por homografías globales hacia el plano de referencia de la primera imagen.



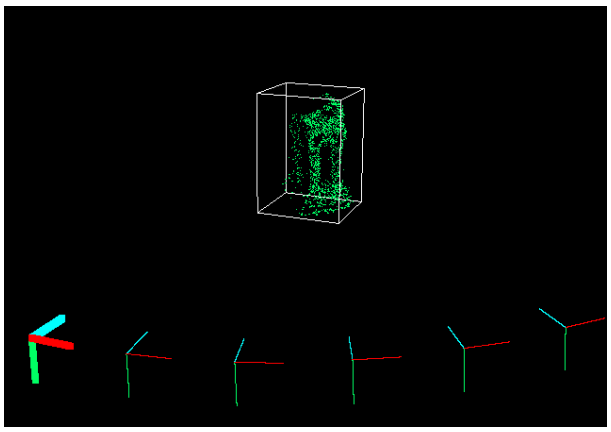
(a) Proyección de máscaras.

(b) Composición por *blending* lineal.

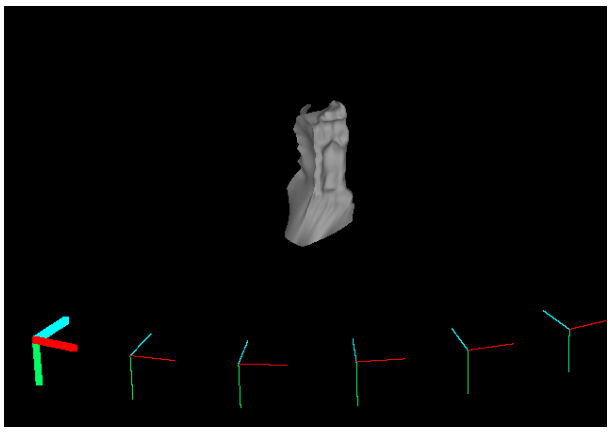
Figura 6.13: Resultado de la composición de la secuencia 'Truck2L' completa, mediante registro por homografías globales hacia el plano de referencia de la primera imagen.

6.4. Resultados adicionales del algoritmo de stitching por reconstrucción de escenas

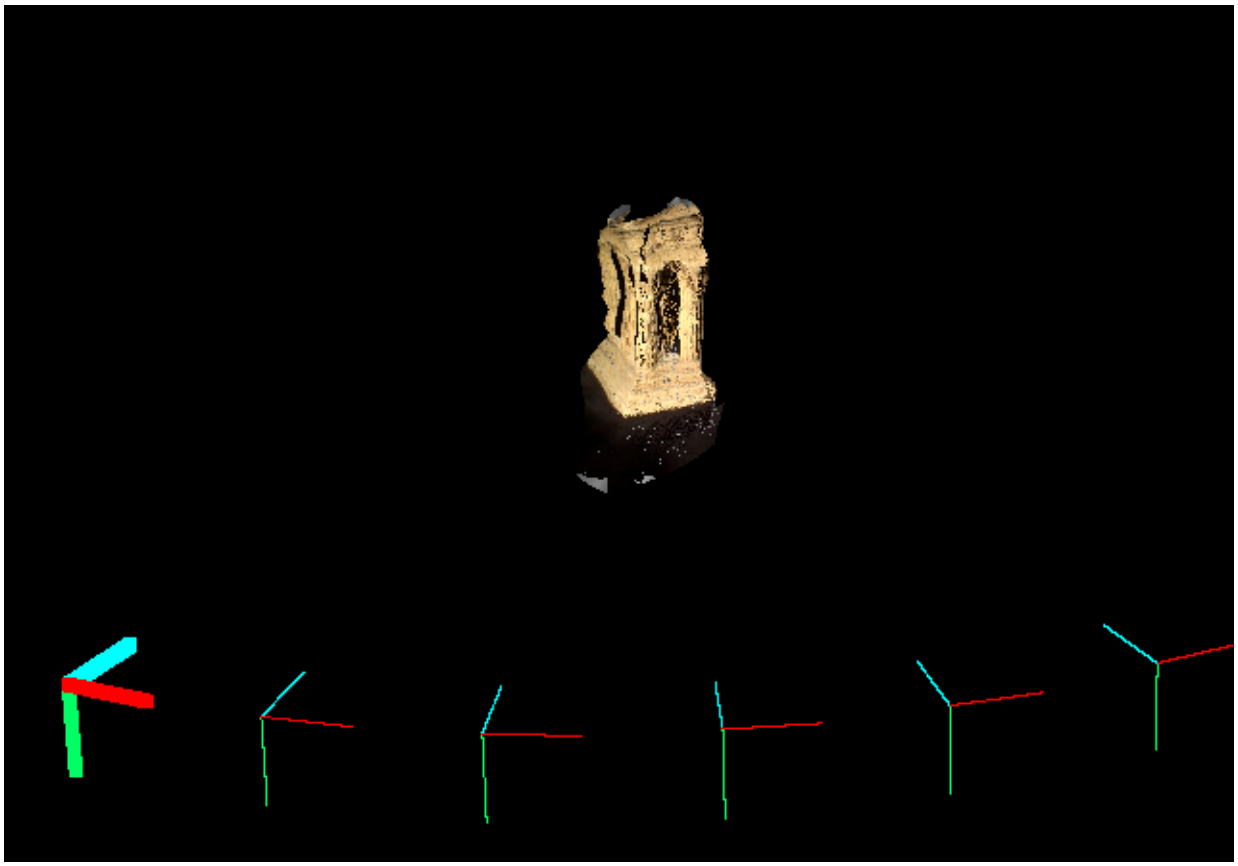
En este anexo, se completan los resultados expuestos en la Sección 4.3.4 para las secuencias restantes, mostrando la nube de puntos, superficie ajustada y composición sobre modelo 3D.



(a) Reconstrucción parcial.

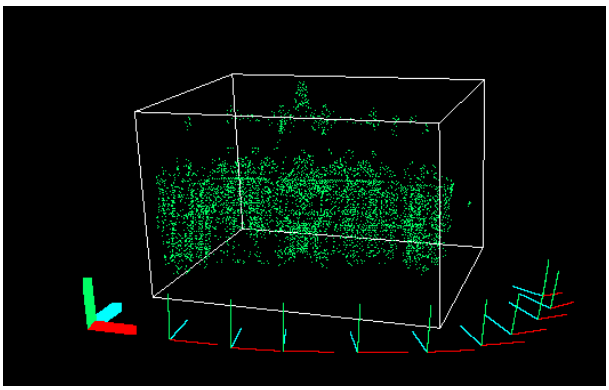


(b) Superficie ajustada para proyección.

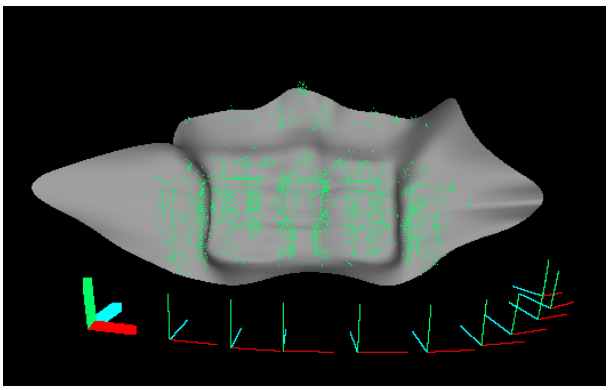


(c) Composición sobre modelo 3D.

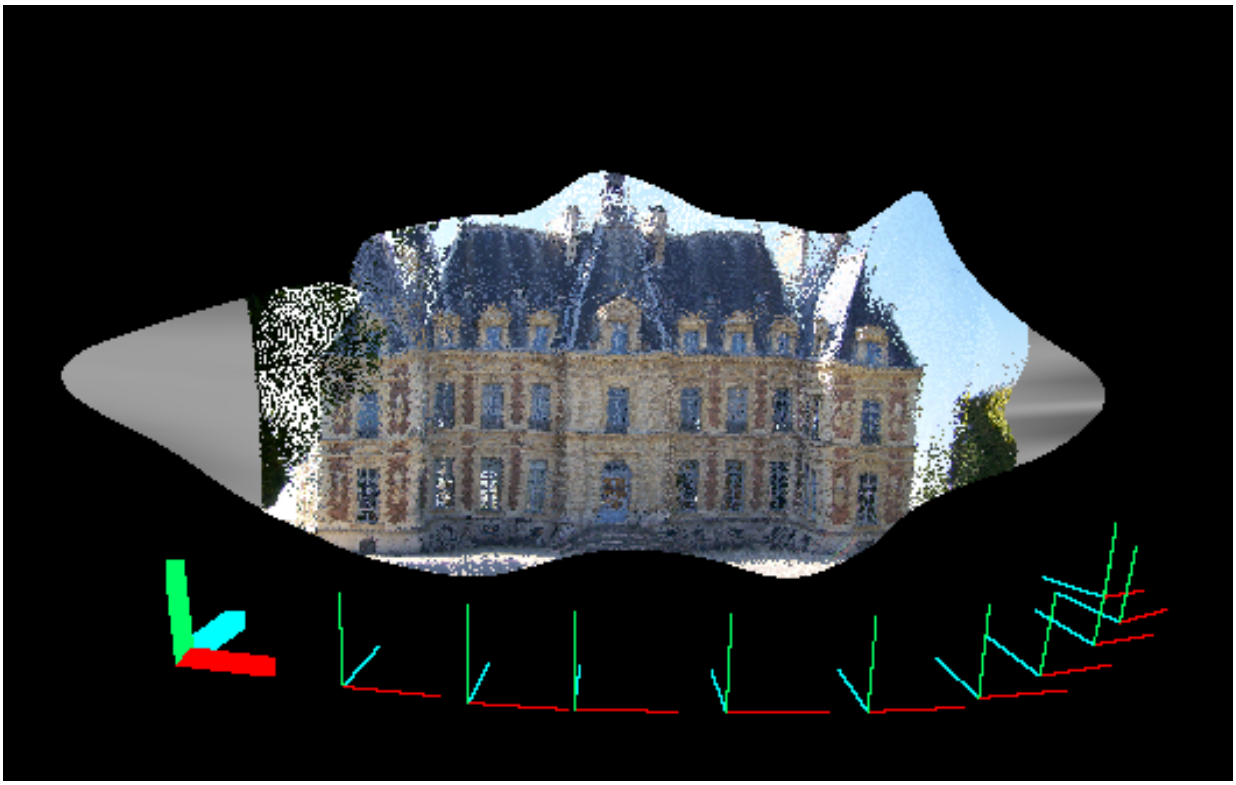
Figura 6.14: Resultado de la reconstrucción de escena sobre la secuencia 'Temple'.



(a) Reconstrucción parcial.

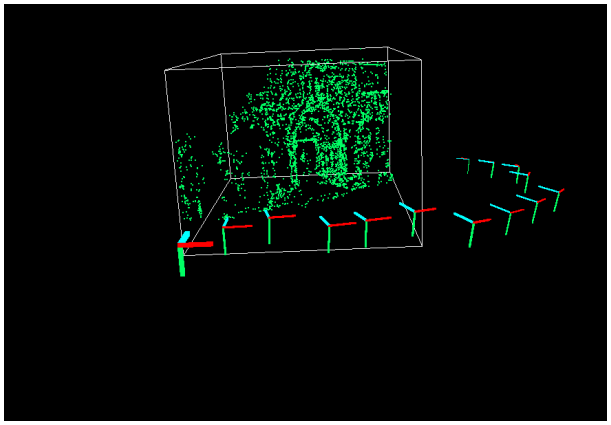


(b) Superficie ajustada para proyección.

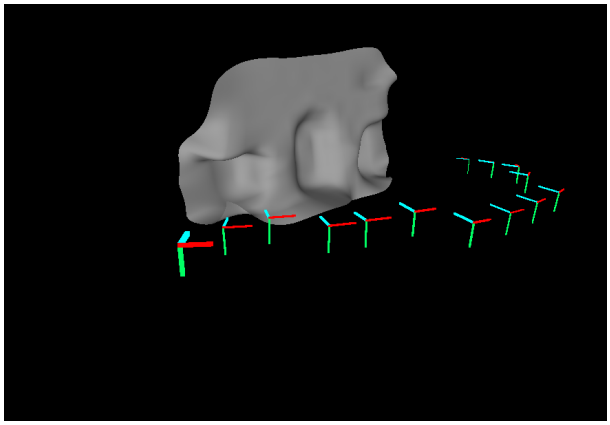


(c) Composición sobre modelo 3D.

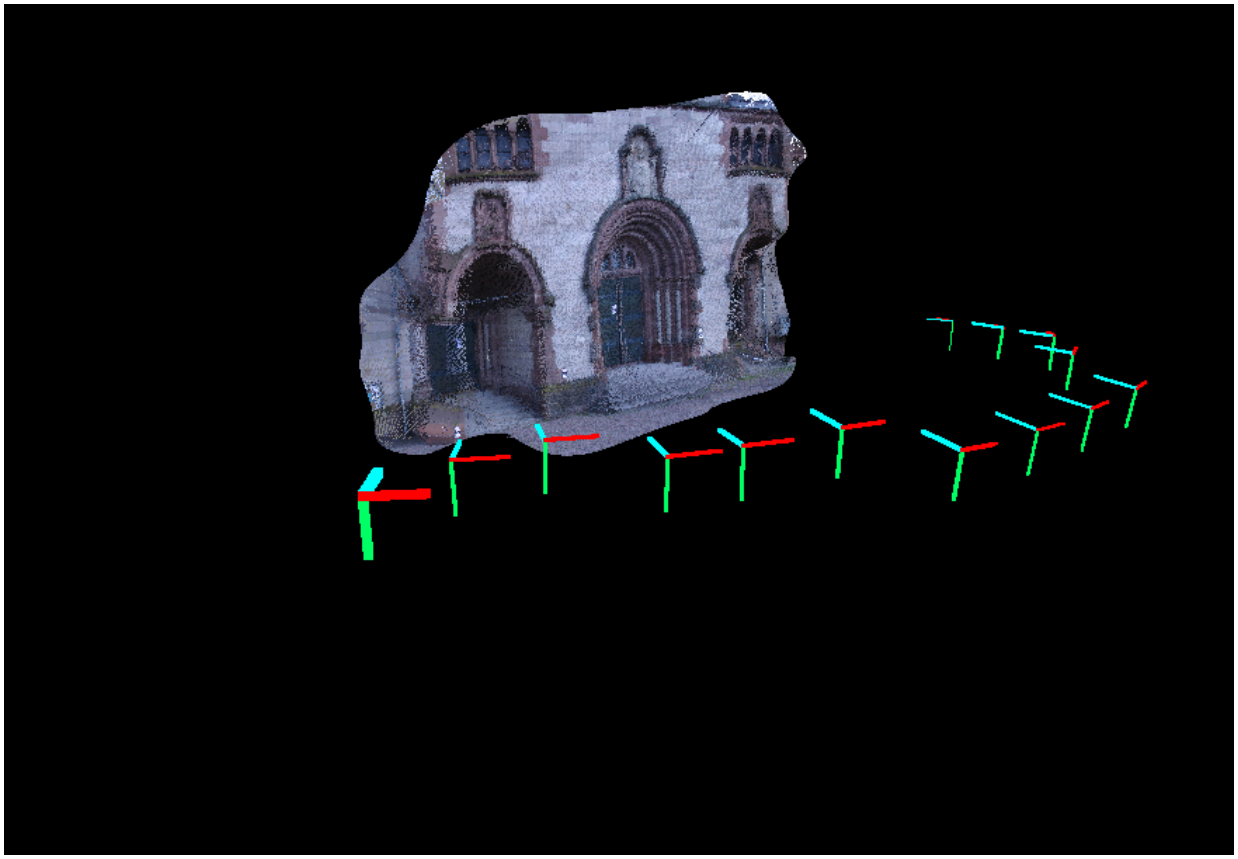
Figura 6.15: Resultado de la reconstrucción de escena sobre la secuencia 'SceauxCastle'.



(a) Reconstrucción parcial.

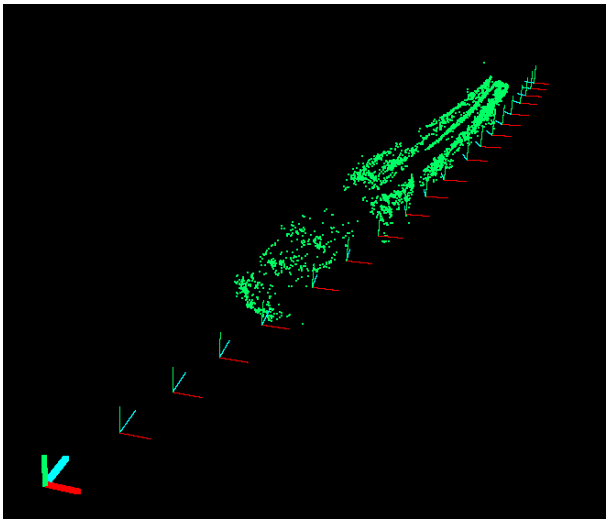


(b) Superficie ajustada para proyección.

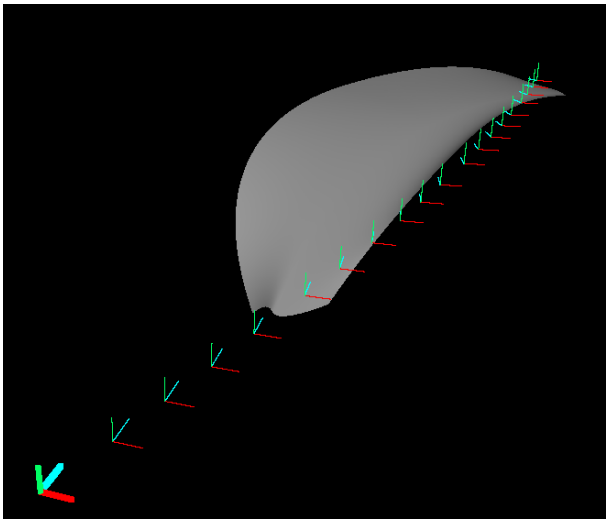


(c) Composición sobre modelo 3D.

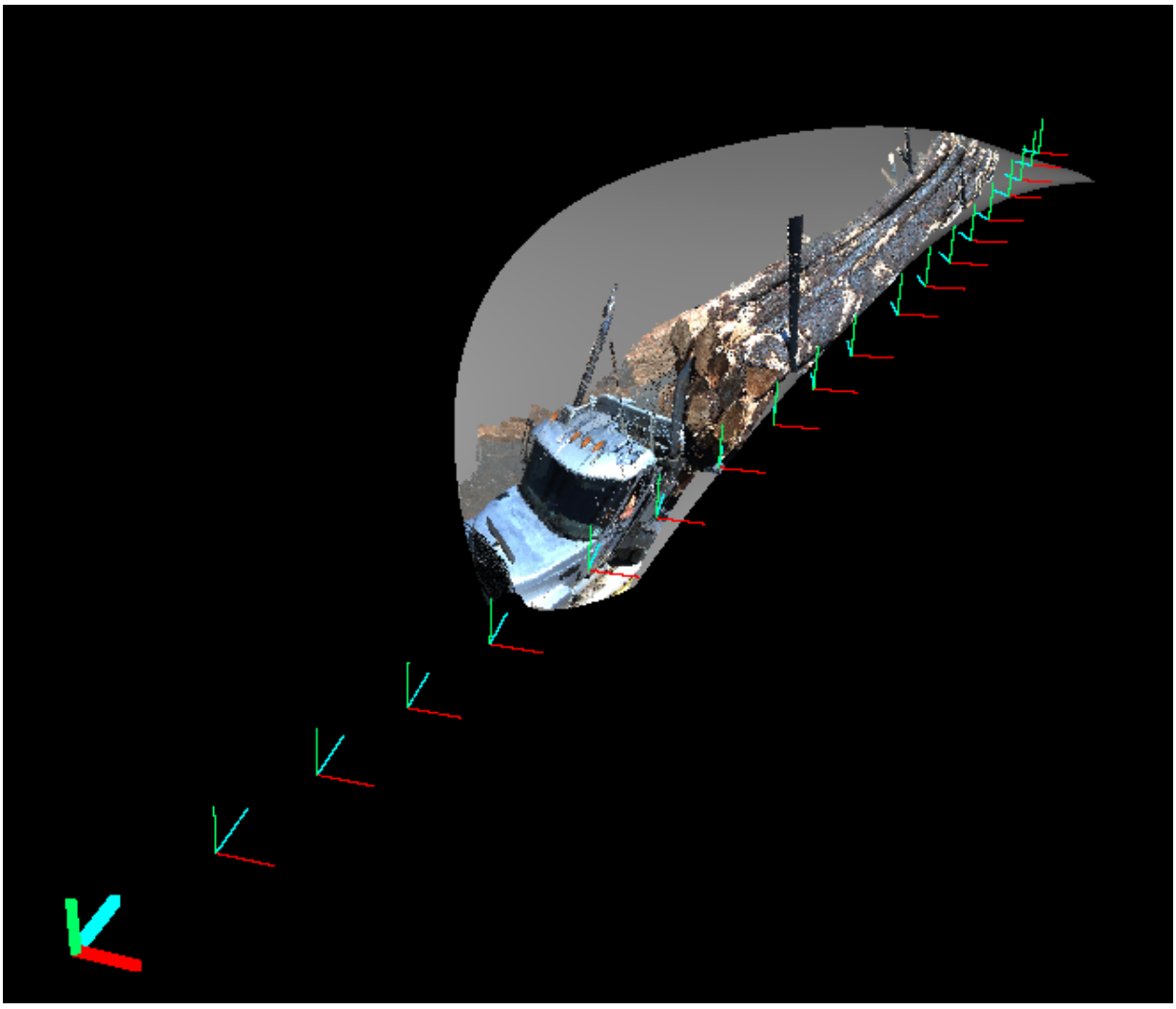
Figura 6.16: Resultado de la reconstrucción de escena sobre la secuencia 'HerzJesu'.



(a) Reconstrucción parcial.

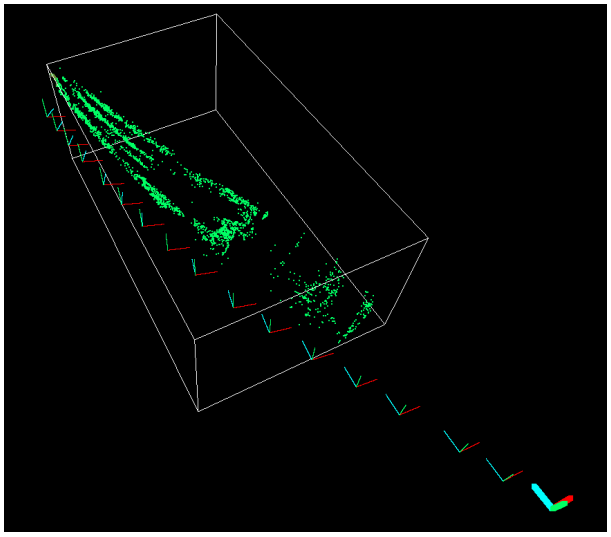


(b) Superficie ajustada para proyección.

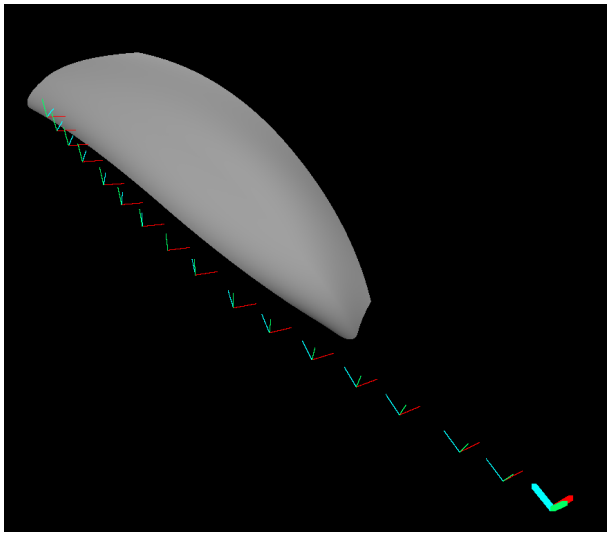


(c) Composición sobre modelo 3D.

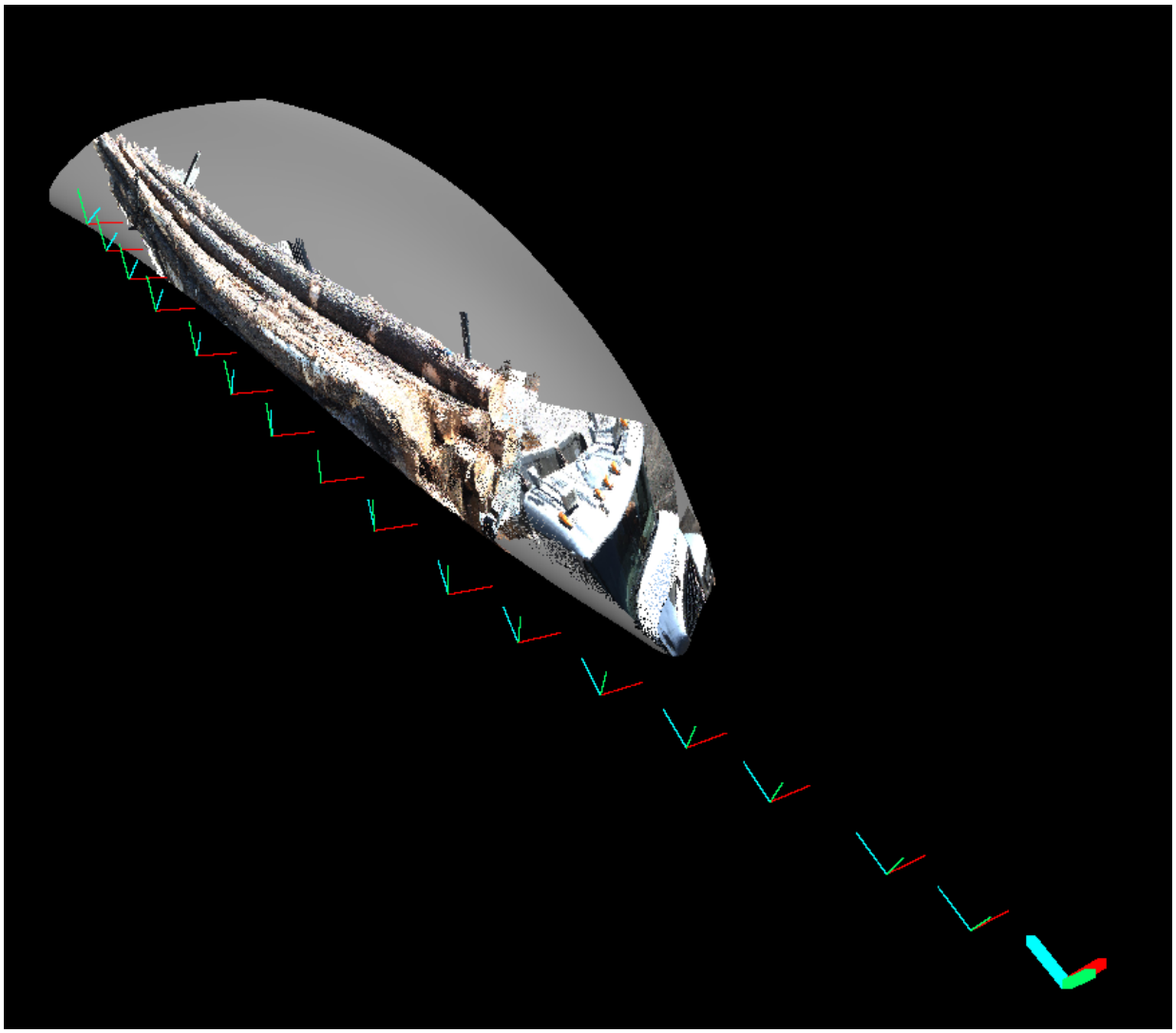
Figura 6.17: Resultado de la reconstrucción de escena sobre la secuencia 'Truck1R'.



(a) Reconstrucción parcial.

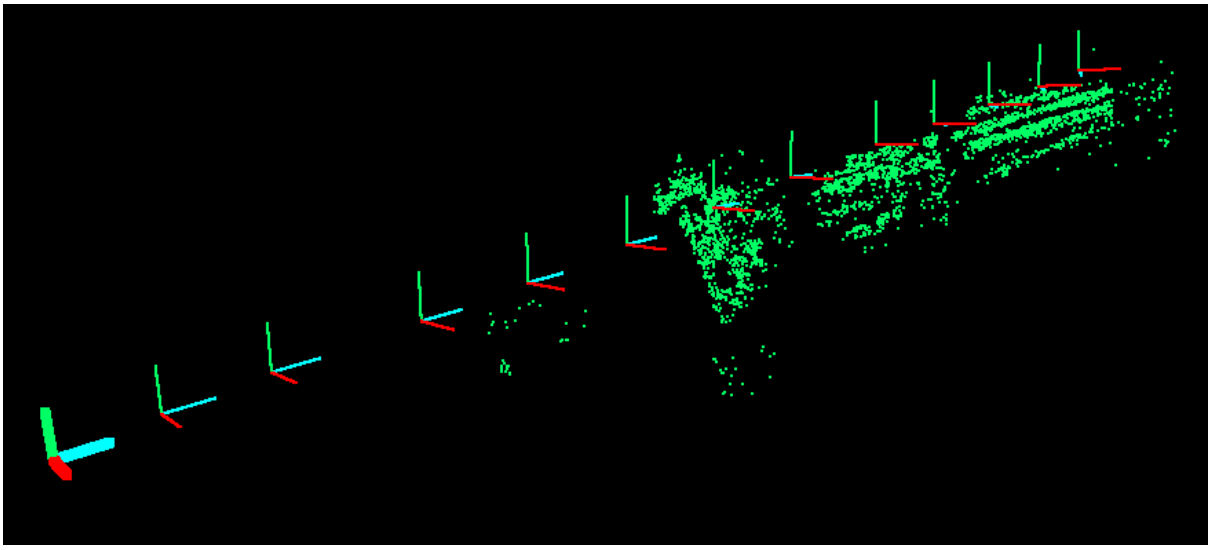


(b) Superficie ajustada para proyección.

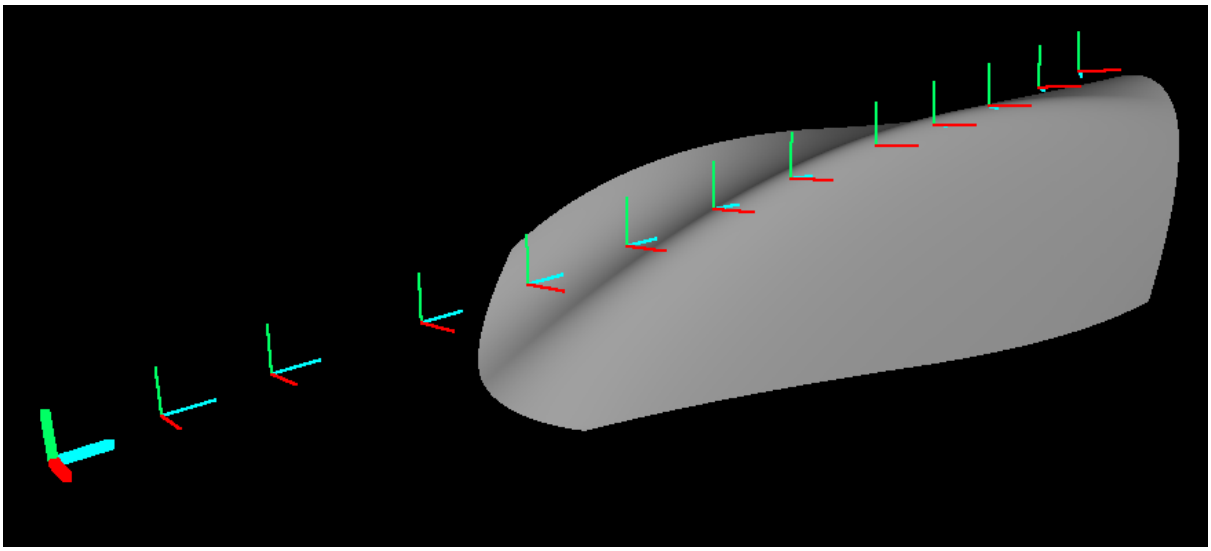


(c) Composición sobre modelo 3D.

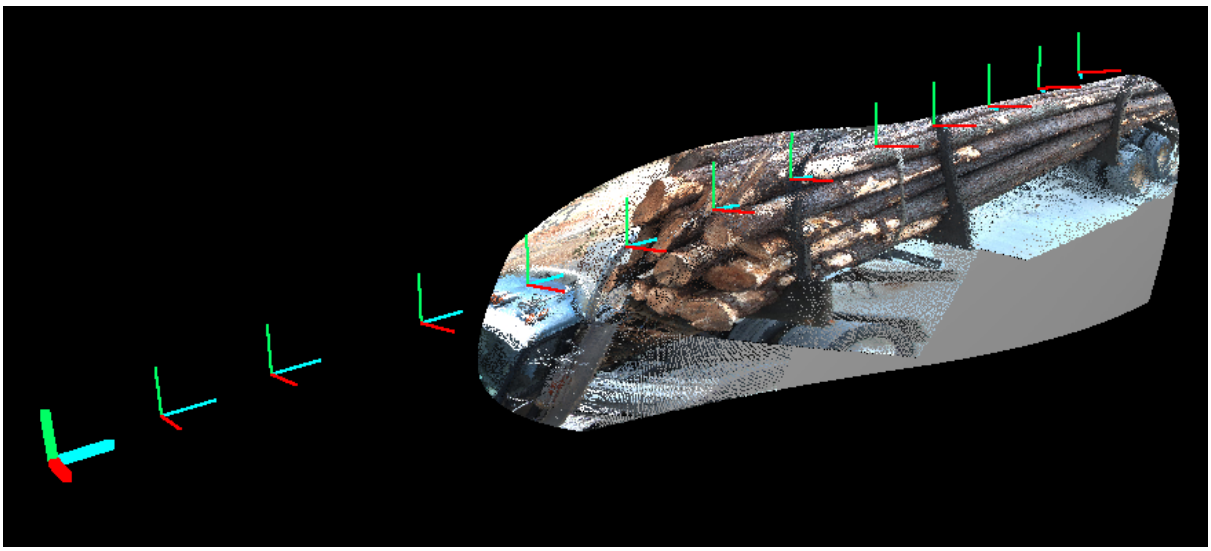
Figura 6.18: Resultado de la reconstrucción de escena sobre la secuencia 'Truck1L'.



(a) Reconstrucción parcial.



(b) Superficie ajustada para proyección.



(c) Composición sobre modelo 3D.

Figura 6.19: Resultado de la reconstrucción de escena sobre la secuencia 'Truck2R'.