



UNIVERSIDAD SIMÓN BOLÍVAR  
DECANATO DE ESTUDIOS PROFESIONALES  
COORDINACIÓN DE INGENIERIA ELECTRÓNICA

**SISTEMA DE GENERACIÓN DE MOSAICOS 2D  
PARA ROBOTS MÓVILES A PARTIR DE VIDEO  
MONOCULAR**

Por:

Victor Yovanni Garcia Carmona

Realizado con la asesoría de:

José de la Cruz Cappelletto Fuentes

**PROYECTO DE GRADO**

Presentado ante la Ilustre Universidad Simón Bolívar  
como requisito parcial para optar al título de  
Ingeniero Electrónico

Sartenejas, Marzo de 2018



UNIVERSIDAD SIMÓN BOLÍVAR  
DECANATO DE ESTUDIOS PROFESIONALES  
COORDINACIÓN DE INGENIERÍA ELECTRÓNICA

ACTA FINAL PROYECTO DE GRADO

**SISTEMA DE GENERACIÓN DE MOSAICOS 2D PARA ROBOTS  
MÓVILES A PARTIR DE VIDEO MONOCULAR**

Presentado por:

**Victor Yovanni Garcia Carmona**

Este Proyecto de Grado ha sido aprobado por el siguiente jurado examinador:

---

José de la Cruz Cappelletto Fuentes

---

Novel Antonio Certad H.

---

Gerardo Fernandez López

**Sartenejas, @día de Mayo de 2018**



UNIVERSIDAD SIMÓN BOLÍVAR  
DECANATO DE ESTUDIOS PROFESIONALES  
COORDINACIÓN DE INGENIERÍA ELECTRÓNICA  
**SISTEMA DE GENERACIÓN DE MOSAICOS 2D PARA ROBOTS  
MÓVILES A PARTIR DE VIDEO MONOCULAR**  
PROYECTO DE GRADO  
PRESENTADO POR:  
Victor Yovanni Garcia Carmona, Carnet: 12-10738

## RESUMEN

Al realizar tareas de exploración para el análisis del suelo en espacios aéreos o en fondo marino, es muy común emplear sistemas de adquisición basados en captura de vídeos para su posterior análisis. En la actualidad, el incremento de la tecnología sobre el procesamiento de datos, ha permitido que los algoritmos de visión por computadora coloquen a la cámara como principal sensor para la reconstrucción de entornos recorridos por vehículos móviles. El presente trabajo se encuentra enfocado al análisis e implementación de distintos algoritmos para la reconstrucción de un mosaico 2D (dos dimensiones) del suelo que recorre un robot, a partir de la información proveniente de una cámara monocular ubicada en la parte inferior de este. El robot en cuestión puede realizar recorridos aéreos para realizar la adquisición del vídeo, o incluso trayectorias mas desafiantes como serian las aplicaciones subacuáticas. Para esto, se implementarán distintos algoritmos usando técnicas de procesamiento de imágenes y visión por computadora, para la elaboración de un sistema automatizado que permita generar un mapa en dos dimensiones de la trayectoria recorrida por el vehículo móvil, mejorando la detección de puntos clave y optimizando el calculo de las matrices de transformación para la alineación de imágenes en el mosaico, logrando así un mapa con la menor distorsión posible. Además de realizar análisis sobre el error de proyección de dichas imágenes en el mapa del suelo generado.

**Palabras clave:** mosaico, video monocular, puntos clave, matriz de transformación.

# Agradecimientos

# Índice general

|   |           |
|---|-----------|
| <b>Resumen</b>  | <b>I</b>  |
| <b>Agradecimientos</b>  | <b>I</b>  |
| <b>Índice de Figuras</b>                                      | <b>IV</b> |
| <b>Lista de Tablas</b>  | <b>V</b>  |
| <b>Acrónimos y Símbolos</b>                                   | <b>VI</b> |
| <br>  |           |
| <b>1. Introducción</b>  | <b>1</b>  |
| 1.1. Antecedentes . . . . .                                   | 1         |
| 1.2. Justificación y planteamiento del problema . . . . .     | 2         |
| 1.3. Objetivos . . . . .                                      | 4         |
| 1.3.1. Objetivo General . . . . .                             | 4         |
| 1.3.2. Objetivos Específicos . . . . .                        | 4         |
| 1.4. Estructura del trabajo . . . . .                         | 5         |
| <br>  |           |
| <b>2. Sistemas de generación de mosaico</b>                   | <b>7</b>  |
| 2.1. Estado del arte . . . . .                                | 8         |
| 2.2. Esquema propuesto . . . . .                              | 16        |
| 2.3. Librería de desarrollo . . . . .                         | 17        |
| 2.3.1. OpenCV . . . . .                                       | 17        |
| <br>  |           |
| <b>3. Detección de puntos característicos</b>                 | <b>19</b> |
| 3.1. Introducción . . . . .                                   | 19        |
| 3.2. Revisión teórica . . . . .                               | 20        |
| 3.2.1. Detectores y descriptores de características . . . . . | 20        |
| 3.2.2. Emparejadores de puntos característicos . . . . .      | 29        |
| 3.3. Módulo comparativo . . . . .                             | 30        |
| 3.3.1. Resultados . . . . .                                   | 30        |
| 3.4. Conclusiones . . . . .                                   | 30        |

---

|   |           |
|---|-----------|
| <b>4. Alineación de imágenes</b>                        | <b>31</b> |
| 4.1. Introducción . . . . .                             | 31        |
| 4.2. Revisión teórica . . . . .                         | 31        |
| 4.2.1. Transformaciones geométricas . . . . .           | 32        |
| 4.3. Generación de sub-mosaicos . . . . .               | 32        |
| 4.3.1. Selección de imagen de referencia . . . . .      | 32        |
| 4.3.2. Matriz de transformación promedio . . . . .      | 33        |
| 4.4. Corrección euclidiana . . . . .                    | 33        |
| 4.5. Resultados . . . . .                               | 33        |
| 4.6. Conclusiones . . . . .                             | 33        |
| <b>5. Unión de imágenes</b>                             | <b>34</b> |
| 5.1. Introducción . . . . .                             | 34        |
| 5.2. Línea de costura . . . . .                         | 34        |
| 5.2.1. Mapa más cercano . . . . .                       | 34        |
| 5.2.2. Corte por grafo . . . . .                        | 34        |
| 5.3. Corrección de color . . . . .                      | 35        |
| 5.3.1. Ajuste de ganancia . . . . .                     | 35        |
| 5.3.2. Método de Reinhard . . . . .                     | 35        |
| 5.4. Fusión de imágenes . . . . .                       | 35        |
| 5.4.1. Fusión ponderada . . . . .                       | 35        |
| 5.4.2. Fusión piramidal . . . . .                       | 35        |
| 5.5. Resultados . . . . .                               | 35        |
| 5.6. Conclusiones . . . . .                             | 35        |
| <b>6. Conclusiones y trabajos futuros</b>               | <b>36</b> |
| <b>A. Instalación de librería OpenCV y dependencias</b> | <b>40</b> |
| A.1. @sección . . . . .                                 | 40        |
| A.1.1. @subsección . . . . .                            | 40        |

# Índice de figuras

|   |    |
|---|----|
| 1.1. Robot móvil OpenROV . . . . .                                      | 2  |
| 1.2. Efectos en el cambio del punto de vista . . . . .                  | 3  |
| 1.3. Logo del software Hugin . . . . .                                  | 3  |
| 2.1. Proceso básico para la generación de mosaico . . . . .             | 9  |
| 2.2. Clasificación de los algoritmos de generación de mosaico . . . . . | 10 |
| 2.3. Esquema propuesto para la construcción del mosaico . . . . .       | 17 |
| 2.4. Logo de la librería OpenCV . . . . .                               | 18 |
| 3.1. Caracterización de regiones en una imagen . . . . .                | 21 |
| 3.2. Ventana de búsqueda para de detección de esquinas . . . . .        | 22 |
| 3.3. Efecto del escalado sobre las esquinas . . . . .                   | 22 |
| 3.4. Detector SIFT . . . . .  | 24 |
| 3.5. Descriptor SIFT . . . . .  | 25 |
| 3.6. Detector y descriptor SURF . . . . .                               | 25 |
| 3.7. Detector y descriptor ORB . . . . .                                | 26 |
| 3.8. Filtro no lineal propuesto por KAZE . . . . .                      | 28 |
| 3.9. Emparejamiento de puntos característicos . . . . .                 | 29 |

# Índice de Tablas



# Acrónimos y Símbolos

## ***Dedicatoria***

*A @personasImportantes, por @razonesDedicatoria.*

# Capítulo 1

## Introducción

La navegación y exploración en áreas de difícil acceso mediante el uso de robots, es una tarea que se ha venido desarrollando en el Grupo de Investigación y Desarrollo en Mecatrónica de la USB (*GIDM*) desde hace mucho tiempo. Donde una de las aplicaciones mas demandadas, es la tarea de reconstruir un mapa 2D de la superficie mapeada por los robots utilizados. En el presente capitulo se pretende introducir los trabajos previos y avances que se han tenido en el desarrollo de este tipo aplicaciones, específicamente en el *GIDM*, que dieron origen y motivación para la realización del proyecto. Además de postular un serie de problemas que el presente trabajo busca solucionar.

### 1.1. Antecedentes

En el *GIDM* se han realizado grandes avances en el desarrollo de equipos y plataformas robóticas para actividades de investigación, exploración e inspección de ambientes no estructurados. Usualmente cuando se opera en este tipo de ambientes, en busca de realizar exploraciones mas eficientes y a mayor escala, se emplean vehículos operados remotamente *ROV* (del inglés: Remotely Operated Vehicles) equipados con cámaras de vídeo. O bien, para el caso de aplicaciones subacuáticas también se suelen utilizar vehículos autónomos submarinos *AUV* (del inglés: Automated Underwater Vehicles), mientras que para exploraciones aéreas se hace uso de vehículos aéreos no tripulados *UAV* (del ingles: Unmanned Aerial Vehicle).

En este sentido, se tienen proyectos como el presentado por *Danilo, D.* [1], cuyo trabajo de grado consistió en el desarrollo en un sistema de operación remota para un robot submarino (*ROV*), con la finalidad de implementarlo en tareas de exploración. Con objetivos similares, *Said, A.* [2] basó su proyecto de grado en la instrumentación y control de un robot cuadricóptero volador(*AUV*).

Adicional a los proyectos antes mencionados, en el el *GIDM* se cuenta con un vehículo submarino llamado OpenROV<sup>1</sup>, el cual es un robot maniobrado remotamente de baja envergadura, diseñado especialmente para operar bajo el agua.

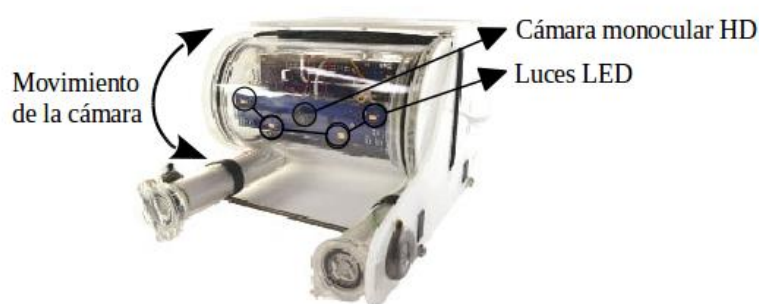


FIGURA 1.1: Robot móvil OpenROV

Si bien se cuenta con un conjunto de plataformas robóticas adaptadas para tareas de exploración, hasta el momento no se han desarrollados sistemas basados en visión que permitan incluirse en la etapa de navegación y mapeo de dichos vehículos, siendo la presente investigación la primera en abordar la tarea de la reconstrucción del suelo recorrido haciendo uso únicamente de una cámara de vídeo, sensor presente en todas las plataformas robóticas antes mencionados.

## 1.2. Justificación y planteamiento del problema

Cuando se habla de construir un mosaico 2D, se hace referencia al proceso de recortar y alinear imágenes, de tal forma que puedan ser representadas todas juntas en una sola gran imagen. Es importante considerar que las imágenes para este tipo de aplicaciones son capturadas desde diferentes ubicaciones de la cámara, a diferencia del proceso para elaborar imágenes panorámicas, en las cuales esta

<sup>1</sup> <https://www.openrov.com/products/openrov28/>

ubicación es una constante. Esta característica trae consigo uno de los principales problemas en la construcción de mosaicos, y se debe al efecto paralaje. Este efecto está asociado a la diferencia entre las posiciones aparentes de los objetos, según el punto desde donde se observa.

En la figura 1.2 se aprecia un ejemplo ilustrativo de esta definición, en la cual, si nos fijamos en el punto de vista **A**, se observa el triángulo a la izquierda del círculo, mientras que desde el punto **B** este orden se encuentra invertido.

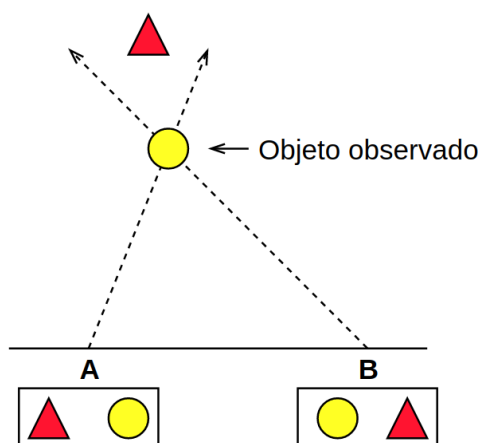


FIGURA 1.2: Efectos en el cambio del punto de vista.

Si bien, este problema afecta en gran medida la construcción del mosaico, no es el único presente, y se intensifican en aplicaciones de mapeo submarino, en las cuales, se evidencian efectos de distorsión de los objetos, absorción y cambios en la dirección de la luz, producto de pequeñas partículas suspendidas en el agua.

En el *GIDM* actualmente se utilizan mecanismos manuales para la elaboración de estos mapas, en específico, se hace uso de *softwares* como Hugin<sup>2</sup>. Este es un programa de código abierto y gratuito bajo licencia GPL<sup>3</sup>, el cual está dedicado a la generación de imágenes panorámicas, incluyendo funciones para el recorte, alineación y corrección de color; además de algoritmos para la optimización de parámetros en la cámara, y corrección de distorsión. Si



FIGURA 1.3: Logo del software Hugin

<sup>2</sup>[hugin.sourceforge.net/](http://hugin.sourceforge.net/)

<sup>3</sup> <http://www.gnu.org/copyleft/gpl.html>

bien este software esta diseñado para la creación de imágenes panorámicas, permite el uso de varios tipos de proyecciones cartográficas, entre estas la rectangular, proyectando las imágenes sobre un plano recto.

Esta practica manual, además de limitar el alcance de los sistemas embebidos para el uso en navegación automática, requiere de una inversión de tiempo importante por medio del usuario en el proceso de selección y alineación de imágenes.

Atendiendo a esta necesidad, es necesario contar con un sistema que permita realizar la reconstrucción del suelo con la menor interacción posible del ser humano. Asimismo, con el fin de poder realizar operaciones de mapeo y localización simultanea *SLAM* (del ingles: Simultaneous Localization and Mapping), haciendo uso de las herramientas y robots existentes en el laboratorio, se requiere contar con un sistema basado en visión, que genere de forma automática un mapa 2D de la superficie sobre la que navega o sobrevuela el vehículo remoto, y que logre lidiar de manera efectiva ante los problemas previamente planteados.

## 1.3. Objetivos

### 1.3.1. Objetivo General

Analizar e implementar un sistema automatizado que permita la reconstrucción de un mapa en dos dimensiones, del suelo recorrido por robot, aéreo o submarino, a través de la información capturada por una cámara ubicada en su parte inferior.

### 1.3.2. Objetivos Específicos

- Análisis comparativo de métodos vigentes en la reconstrucción de mosaicos 2D, a partir de imágenes y videos de entrada.
- Implementación de modulo de pre-procesamiento y corrección de entrada.
- Análisis comparativo de métodos de detección y descripción de puntos característicos.

- Implementación de módulo de alineación de imágenes mediante la detección de puntos característicos.
- Cuantificar el error de proyección y distorsión en los modelos 2D generados.

## 1.4. Estructura del trabajo

Luego de presentar el planteamiento del problema y la descripción del proyecto, la presente investigación se encuentra dividida en 5 capítulos, organizados de la siguiente manera:

En el **Capítulo 2** se presenta una revisión del estado del arte sobre los algoritmos de generación de mosaico, en el cual se exponen los trabajos recientes y avances importantes en esta área de investigación. Al mismo tiempo, se describen los módulos principales que componen este tipo de sistemas. Luego, en base a los algoritmos y técnicas estudiadas se propone un esquema para un sistemas de generación de mosaico. Para finalizar, se presenta la librería de procesamiento de imágenes que se planteó utilizar para la implementación de los algoritmos propuestos.

El **Capítulo 3** inicia una revisión teórica en la cual se describe el funcionamiento de los algoritmos detectores, descriptores y emparejadores de características; y posteriormente se presentan resultados de pruebas comparativas entre los mas usados para este tipo de aplicaciones.

El módulo encargado de la alineación de imágenes en el mosaico, es descrito en el **Capítulo 4**. Al igual que el capítulo anterior, se presenta una revisión teórica de los conceptos necesarios para su implementación. Luego, se introduce el modelo de sub mosaicos, y la implementación de un conjunto de correcciones geométricas sobre este nuevo modelo. Finalmente se muestran los resultados de los algoritmos aplicados en esta sección, seguidos de sus respectivos análisis.

En el **Capítulo 5** se describe el modulo final del sistema, en donde se explica el funcionamiento de los algoritmos que corrigen visualmente el mosaico final. De

igual forma se muestran los resultados de su implementación, seguidos de una conclusión final sobre estos.

Finalmente, en el ***Capítulo 6*** se presentan las conclusiones finales, además de propuestas sobre recomendaciones y posibles implementaciones que pueden aportar mejoras y/o permitir la continuación del proyecto aquí planteado.



## Capítulo 2

# Sistemas de generación de mosaico

En este capítulo se presenta una revisión teórica del estado actual de las investigaciones que se han desarrollado en el área de procesamiento de imágenes, aplicado a la construcción de mosaicos, además de una reseña histórica de la evolución de dichos métodos. Debido a que la construcción de mosaicos ha sido y sigue siendo un área de investigación muy activa, existe una gran variedad de métodos y técnicas que se han empleado para este fin. Con el objetivo de recopilar esta información de manera estructurada, se presenta una clasificación de estos algoritmos en base al modo de abordar los módulos principales para la elaboración de estos mapas.

Con esto se pretende recuperar y trascender el conocimiento acumulado en esta área de estudio, además de familiarizar al lector con los conceptos básicos, necesarios para la comprensión del presente trabajo. Seguidamente se presenta el modelo del sistema de generación de mosaico propuesto en base a los algoritmos y técnicas que se han utilizado en las investigaciones más recientes. Finalmente se introduce la librería de procesamiento de imágenes que se seleccionó para la implementación de todos los módulos necesarios.

## 2.1. Estado del arte

La elaboración de mosaicos para la construcción de mapas del suelo, se ha desarrollado incluso antes desde la era digital de la computadoras. Desde que el proceso de registrar fotografías ha existido, se comenzaron a usar para elaborar mapas topográficos [3], donde imágenes adquiridas a partir de globos aerostáticos o altas colinas eran unidas manualmente. Posteriormente, producto de los avances en materia de aeronáutica, el interés por la aerofotografía se incrementó en gran medida. En este mismo sentido se utilizaban aviones para el registro de imágenes a mayores altitudes, logrando cubrir grandes áreas en menor cantidad de tiempo. Pero debido a que no se alcanzaba suficiente altura, y se mantenía la necesidad de registrar grandes áreas, era requerido que los mapas se construyan mediante fotografías que se superpongan, de igual forma esta tarea se llevaba a cabo mediante técnicas manuales por medio de expertos.

La necesidad de registrar áreas aun mas grandes siguió avanzando, motivado por la llegada de los satélites que eran capaces de enviar a tierra la información que obtenían de las cámaras. Los avances tecnológicos en materia de computación, y el creciente aumento de datos para esta aplicación, promovieron el desarrollo de técnicas de procesamiento digital de imágenes para dar solución a este tipo de problemas.

Con el desarrollo de cámaras cada vez mas pequeñas y portátiles, así como también la llegada de vehículos no tripulados mas compactos —ROV, UAV, AUV—, se produjeron grandes avances y nuevas técnicas por parte de centros de investigación en el área de la física, robótica y visión por computadora, que buscaron aportar soluciones para la realización automática de mosaicos. Esta vez, con un creciente enfoque en las aplicaciones mas desafiantes como los ambientes submarinos.

Para explicar el proceso en el que consiste construir un mapa del suelo a partir de fotografías, lo podemos separar en tres simples pasos. Si bien, éste se ilustran gráficamente en la figura 2.1, a continuación se explica cada una de estas etapas.

- **Registro:** De su termino en inglés *image-registration*, consiste en establecer la correspondencia geomántica entre las imágenes que componen la misma

escena. Para esto, es necesario estimar la transformación geométrica que logra alinear dichas imágenes en un mismo plano.

- **Alineación:** También llamada proyección, consiste en alinear las imágenes registradas en un sistema de referencia común, es decir, con respecto a un plano de referencia. En este caso se utiliza la transformación geométrica calculada en el paso anterior.
- **Fusión:** En este paso se busca corregir los errores fotométricos o discontinuidades, presentes en el mosaico luego del proceso de alineación. Estos errores aparecen, producto de errores en la estimación de las transformaciones o a cambios en la perspectiva de los objetos observados.

Si bien, se han propuesto una gran cantidad de algoritmos por parte de distintos grupos de investigación en todo el mundo, esta tarea aun sigue siendo desafiante, debido mayormente a los procesos de registro y fusión de las imágenes.

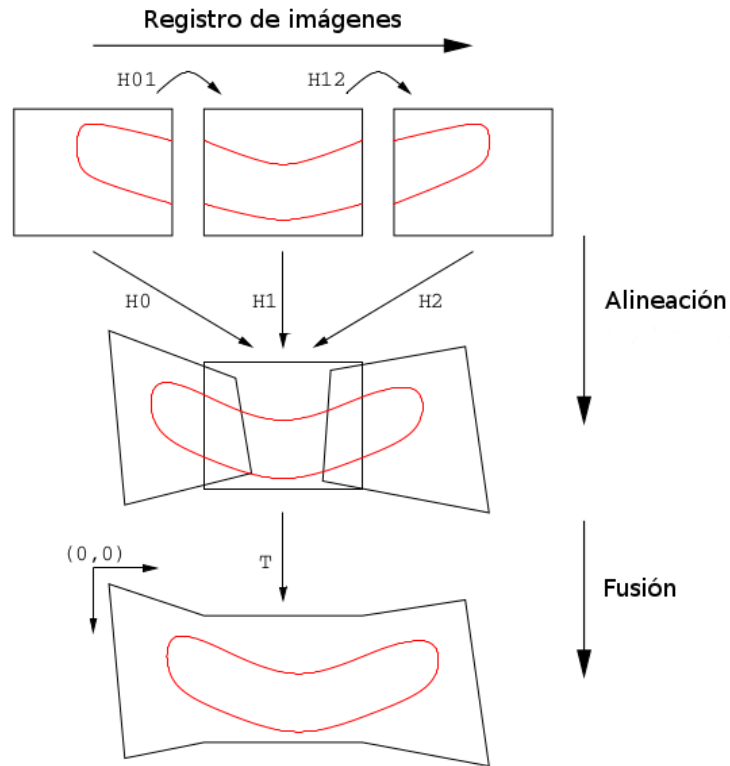


FIGURA 2.1: Proceso básico para la generación de mosaico, adaptado de [4]

Específicamente la etapa para estimar correspondencias entre las imágenes es un problema complicado, en principio debido a la naturaleza no plana de los suelos

estudiados; por otro lado la reducción de discontinuidades, o inconsistencias entre imágenes consecutivas sigue siendo una tarea desafiante, en este caso debido a fuertes cambios de exposición o errores en la etapa de registro. Es por esto que la mayoría de avances e implementaciones en esta área, están encaminados en resolver estos dos problemas principales, o bien mejorar los resultados de trabajos previos.

De esta forma se propone una clasificación de los algoritmos de mosaicos, basada en como estos abordan los procesos de registro y fusión. Además, para cada clasificación se realiza una breve revisión teórica de cada categoría, así como también los diferentes métodos y modificaciones que han aplicado diversos desarrolladores. en la figura 2.2 se puede apreciar esta clasificación.

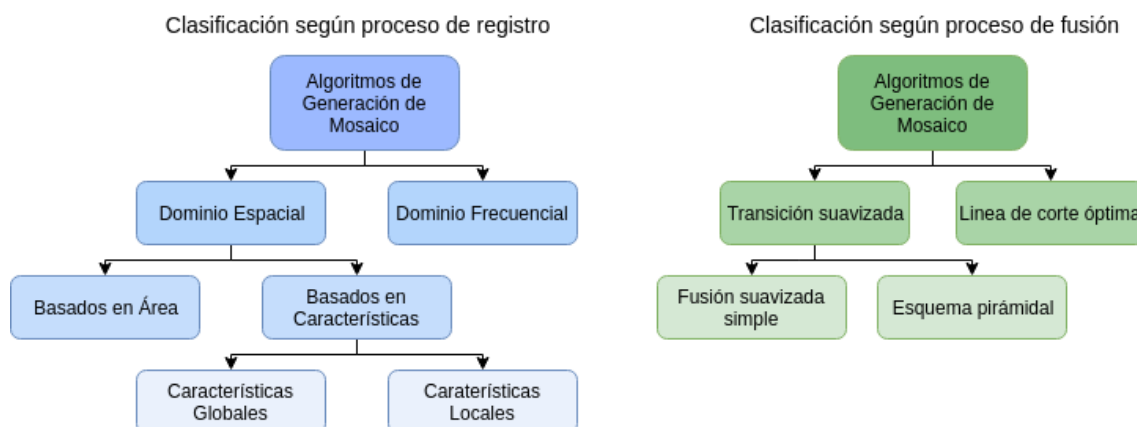


FIGURA 2.2: Clasificación de los algoritmos de generación de mosaico según como aborden las etapas de registro y fusión

## Clasificación basada en el registro de imágenes

Este proceso es muy importante para la creación de mosaicos, y básicamente es la base para estos. Cuando se registran imágenes, primero lo que se busca es encontrar la relación, o la correspondencia entre estas, teniendo en cuenta que pudieron haber sido capturadas desde distintos puntos de vista, distintos instantes de tiempo, distinta perspectiva, o incluso distintas cámaras. Luego de encontrar las zonas o puntos correspondientes, se busca estimar una matriz de transformación geométrica que permita alinearlas todas en un sistema de referencia común. Se

puede decir que el registro ha sido exitoso, si se logra estimar una matriz de transformación tal que todos los puntos correspondientes se puedan unir.

Las relaciones entre las imágenes se pueden establecer utilizando distintos métodos, ya sea, emparejando puntos coincidentes, regiones enteras, o bien usando la propiedad de correlación de fase en el dominio de la frecuencia. Si bien se cuenta con algoritmos de generación de mosaico que realizan el registro de imágenes basados en datos de navegación, como el GPS (del inglés: Global Positioning System), o instrumentos de medición inercial IMU (del inglés: Inertial Measurement Unit), para estimar el movimiento de los vehículos y así establecer una relación entre las imágenes basado en el movimiento de la cámara, en este trabajo solo se discutirán técnicas que utilicen únicamente una cámara monocular como sensor de entrada. Dicho esto, estos métodos para establecer correspondencias son discutidos a continuación.

## **Algoritmos en el dominio espacial**

Los algoritmos en esta categoría utilizan la información de los píxeles para establecer la relación entre imágenes, es decir, se utiliza el valor de los píxeles (intensidad) y se trata de establecer la correspondencia de estos según la ubicación en la que se encuentran. Estos se pueden separar en dos técnicas principales: basados en área o en características.

Los algoritmos basados en área, buscan relacionar dos ventanas o regiones en dos imágenes que correspondan a la misma escena. El concepto principal consiste en mover la región de interés desde la primera imagen hacia la segunda, buscando que la diferencia entre las intensidades sea la menor posible, es decir, se trata de estimar la mejor matriz de transformación que logre reducir la diferencia de intensidades al alinear las regiones estudiadas [5]. Algunos trabajos importantes en esta clasificación utilizaron algoritmos como NCC (siglas del inglés: Normalized Cross Correlation) [6], y el MI (siglas del inglés: Mutual Information) [7], donde estos proporcionan una métrica de igualdad entre dos imágenes.

El primer caso, el NCC mide la similitud entre las regiones estudiadas según los valores de intensidades, mientras que el MI la mide en base a la cantidad de

información que comparten estas imágenes en términos de entropía. Al emplear esta técnica se logra emparejar las imágenes a nivel de píxel. Si bien se logran buenos resultados, este tipo de algoritmo requiere un alto nivel de superposición entre las imágenes de entrada. Además, el proceso de iterar para optimizar los parámetros de transformación y el calculo del error para cada píxel sobre las regiones, se convierte computacionalmente costoso.

Para reducir el tiempo de computo, se utilizan algoritmos basados en la relación de características, en los cuales se tratan de detectar puntos o regiones en distintas imágenes, que correspondan con la misma ubicación. Estas características detectadas se pueden evidenciar en forma de puntos aislados, curvas continuas o regiones conectadas. Luego se puede encontrar la transformación geométrica que relaciona las características de origen con las de destino, en muchos casos resolviendo un sistema de ecuaciones. En este caso, el proceso de registro así como también el resultado del mosaico, será tan bueno como el algoritmo de detección que se utilice.

Tal y como se mencionaron los tipos de características que se pueden extraer, podemos clasificar de forma general los algoritmos de detección. Ya sea si trabajan con características locales, como lo son aquellos que detectan puntos aislados. o mas globales como los basados en detección de contornos.

### **Detectores locales**

Al usar este tipo de algoritmos se busca encontrar la relación entre una serie de puntos dispersos que se corresponden entre dos imágenes, donde las características locales mas comunes que se suelen detectar serian esquinas, bordes, manchas, entre otros. Posteriormente el proceso de registro se completa al estimar la transformación geométrica con dicha relación de puntos, en este caso resolviendo un sistema de ecuaciones.

Una de las ventajas principales de esta técnica para la generación de mosaicos, es que puede trabajar con imágenes consecutivas que no posean un alto nivel de cobertura. Siempre y cuando la cantidad de puntos detectados, y correctamente emparejados entre el par de imágenes supere el mínimo necesario para la solución del sistema de ecuaciones.

Diversos algoritmos detectores de características locales se han venido desarrollando desde hace mucho tiempo [8–15], y en la actualidad estos avances han permitido que el uso de este método traiga consigo muchas ventajas sobre el resto, desde variedad de aplicaciones, robustez ante distinto tipo de escenas, y velocidad de computo (siempre en función del detector a utilizar), lo que lo convierte en uno de los mas usados para la construcción de mosaicos.

### **Detectores globales**

Al utilizar este tipo de detectores, se busca encontrar formas, contornos, texturas, o regiones sobresalientes que se mantengan invariantes ante cambios del punto de vista o iluminación [16]. Al igual que con los detectores locales, aquí se busca extraer tanto la posición, como el tamaño y la orientación de estas regiones.

El resto del proceso para completar la etapa de registro se mantiene igual con este método, donde la transformación geométrica se obtiene a partir de la correspondencia entre las posiciones y orientaciones de las regiones de interés que se lograron extraer. Si bien tienen buen rendimiento ante cambios de movimiento desafiantes, su uso implica un aumento en el tiempo de computo.

### **Algoritmos en el dominio frecuencial**

Ya se ha visto que los algoritmos que operan en el dominio espacial cubren la mayoría de las aplicaciones e investigaciones. Sin embargo se pueden encontrar métodos que obtienen los parámetros óptimos para la transformación a partir de cálculos en el dominio de la frecuencia.

Estos algoritmos utilizan la propiedad de correlación de fase para lograr su objetivo. Teniendo un par de imágenes que se encuentran relacionadas por una simple traslación, el funcionamiento consiste en calcular la correspondiente transformada de *Fourier*, luego el espectro de la potencia cruzada entre ambas. De aquí, se asegura que la fase del espectro de la potencia cruzada corresponde con la diferencia de traslación entre las dos imágenes. Finalmente el proceso continua similarmente

a los métodos anteriores, alineando las imágenes según la transformación obtenida, seguido del proceso de fusionarlas.

Tal y como se explicó este proceso para una traslación, se tienen diversos trabajos como [17] que añaden modificaciones para permitir otro tipo de transformaciones como la rotación, e incluso otros que admiten cambios en la escala [18]. Si bien, los trabajos mencionados presentaron importantes, se requiere de un buen porcentaje de cobertura entre las imágenes, y además presentan limitaciones en los grados de libertad de las transformaciones geométricas que se pueden estimar.

## Clasificación basada en la fusión de imágenes

Si bien el proceso de registro de imágenes es fundamental para lograr un mosaico correcto, el paso final de unir las imágenes también es de gran importancia. Teniendo en cuenta que se busca aparentar que todas las imágenes componen una sola, es vital que se logre un mapa final sin inconsistencias o discontinuidades producto de los cambios en iluminación, objetos en movimiento, entre otros.

Debido a la importancia de este paso, numerosos métodos para lidiar con este tipo de problemas se han desarrollado, lo que nos permite también clasificar los algoritmos de generación de mosaico según como aborden este problema. Entre los métodos mas utilizados encontramos: fusionar las imágenes mediante cambios suavizados o búsqueda de la mejor línea de corte.

### Transición suavizada

Los algoritmos de esta categoría tratan de minimizar la diferencia entre dos imágenes, buscando que el cambio entre los bordes de estas sea imperceptible. El método mas simple para fusionar dos imágenes bajo esta técnica, consiste en realizar una suma ponderada sobre el área de superposición entre ambas, ponderando la intensidad de cada imagen a la mitad. Al realizar esta operación se suelen tener efectos indeseados como el efecto fantasma, en el cual se pueden observar duplicados del mismo objeto con cierto nivel de desvanecimiento. Esto se debe a errores



en el proceso de alineación de las imágenes, diferencia en la iluminación, o incluso a objetos móviles.

Para evitar esto, se utiliza un proceso de fusión que consiste en realizar una suma ponderada entre ambas imágenes, pero dando mayor peso a las regiones que se encuentren mas cerca del centro de la imagen, y menor a aquellas que se encuentren cerca el borde. Si bien, se logra reducir posibles discontinuidades entre los bordes originales, el efecto fantasma aun se puede apreciar para imágenes con fuertes problemas de alineación.

Considerando este problema, y con el objetivo de realizar una unión mas robusta, se desarrolló un esquema piramidal de fusión ponderada. El proceso consiste en obtener una imagen laplaciana para distintos tamaños de escala, formando así una pirámide, al mismo tiempo se va creando para cada nivel de la pirámide una mascara difuminada por el efecto del filtro gaussiano de dicha escala. Luego para cada nivel de la pirámide se aplica el algoritmo de fusión ponderada descrito previamente donde la mascara difuminada pondera el valor de cada píxel. Entre los trabajos que aplican este algoritmo obteniendo resultados notables se tiene [19], logrando reducir en gran medida el efecto duplicado en las regiones de superposición.

## **Linea de corte óptima**

En lugar de buscar reducir las posibles discontinuidades a partir de una suave transición a través del borde entre dos imágenes, tal y como lo hacen los algoritmo previos, en este tipo de algoritmos se busca modificar este borde. Es decir, se trata de encontrar la linea de corte en el área de superposición que logre reducir la discontinuidad de texturas entre ambas imágenes. La diferencia principal de este método sobre los anteriores, es que se toma en cuenta la información presente en la región que se desea fusionar, permitiendo que se logre remover errores producidos por el efecto paralaje, o debido a objetos móviles dentro de la escena. Por otra parte, como las regiones resultantes luego de la linea de corte no se comparten información, se pueden presentar discontinuidades producto de grandes diferencias de iluminación.

En esta sección podemos destacar el trabajo en [20], que presenta resultados

robustos ante imágenes con alto efecto paralaje. En este se busca el área para el cual ambas imágenes presentan mejor similitud —llamada área de costura-local—, y se busca una línea de corte que bordeé esta región. Por otro lado tenemos trabajos como el presentado en [21], donde se busca la línea de corte que minimiza la diferencia entre los gradientes en el área de superposición, a través de un modelado por grafos, y resolviendo el algoritmo flujo-máximo/mínimo-corte [22] para encontrar la línea deseada.

## 2.2. Esquema propuesto

En base a los métodos estudiados en la sección anterior, se plantea el siguiente esquema para la construcción automática de mosaicos:

En primer lugar, se propone utilizar un esquema basado en características para la etapa de registro de imágenes, específicamente, utilizar algoritmos detectores de características locales. El uso de esta técnica permite que se pueda trabajar con imágenes consecutivas que no tengan un elevado nivel de superposición. Además, la gran variedad de algoritmos presentes de esta categoría permite trabajar bajo un gran número de ambientes, presentando robustez ante variados tipos de condiciones.

Por otro lado, evaluando visualmente los resultados obtenidos en las distintas técnicas de fusión, se pretende implementar una combinación en los algoritmos que fueron mas eficientes bajo condiciones de alto efecto paralaje, presencia de objetos móviles, cambios en la iluminación, entre otros. En este caso, se buscaría la línea de corte óptima y luego se aplica una fusión bajo un esquema piramidal.

Si bien el registro y la fusión constituyen las etapas mas importantes, diversos métodos para lograr corregir errores de distorsión y proyección se han desarrollado de la mano de muchos grupos de investigación. Evaluando entre los presentes, se plantea implementar el modelo de sub mosaicos propuesto por *F. Bellavia et. al.* en [23, 24], el cual se presentará detalladamente en el siguiente capítulo. Explicado brevemente, estos algoritmos buscan reducir errores provenientes de la sección de registro, y toman en consideración algunos aspectos importantes que logran reducir distorsiones geométricas en el mosaico final.

Finalmente, tomando en cuenta algunas observaciones y recomendaciones de los trabajos consultados, se propone implementar un módulo de corrección de color con el objetivo de reducir las diferencias de intensidades entre los bordes de las imágenes consecutivas. Adicionalmente en el esquema se agregan algunas técnicas propuestas en [25] para añadir robustez al sistema propuesto. El esquema propuesto se puede observar en la figura 2.3, donde cada uno de estos módulos serán explicados en detalle en los capítulos siguientes.

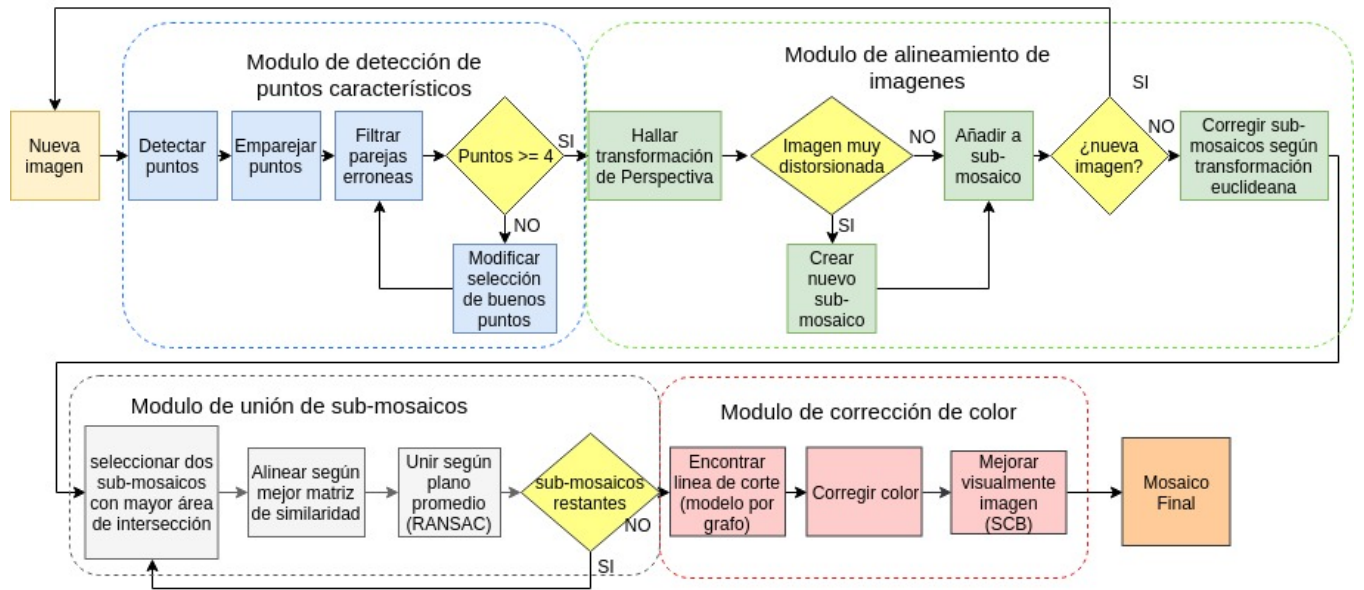


FIGURA 2.3: Esquema propuesto para la construcción del mosaico

## 2.3. Librería de desarrollo

### 2.3.1. OpenCV

Para la implementación de los módulos previamente descritos, es necesario el uso de librerías y entornos de trabajo, que permitan un manejo eficiente de las imágenes. Además, el uso de plataformas que se encuentren estandarizadas en esta área de estudio, facilita que el desarrollo del presente trabajo siga avanzando de la mano de futuros desarrolladores. En base a esto, se seleccionó la librería OpenCV<sup>1</sup> para la implementación de los módulos necesarios en el sistema de generación de mosaico.

<sup>1</sup><http://opencv.org/>

OpenCV (del inglés: *Open Source Computer Vision*) es una librería de procesamiento de imágenes desarrollada por la empresa Intel<sup>2</sup> en el año 1999. Esta librería ofrece un gran numero de algoritmos optimizados (actualmente mas de 2.500), el cual proporciona un entorno de desarrollo altamente eficiente para aplicaciones de procesamiento de imágenes. Asimismo presenta una gran aceptación por parte de los usuarios en el mundo académico y comercial, con mas de 47 mil usuarios activos, y un numero de descargas que supera los 14 millones.



FIGURA 2.4: Logo de la librería OpenCV

Esta plataforma tiene soporte para distintos sistemas operativos, como *Windows*, *Linux*, *Mac OS*, *iOS* y *Android*. Además de tener la posibilidad de trabajarla con diversos lenguajes de programación como: C++, Python, JavaScript. La motivación de el presente trabajo se encuentra orientada al desarrollo de una aplicación que en un futuro pueda ser embebida en un sistema de navegación automático, con lo cual el soporte de un lenguaje de bajo nivel como C++, puede permitir el desarrollo de un algoritmo con suficiente velocidad de computo para este fin.

Por otra parte, esta librería presenta soporte para trabajar con la arquitectura de de cálculo paralelo *CUDA* (del inglés: *Compute Unified Device Architecture*) de la empresa NVIDIA<sup>3</sup>, con la cual se puede aprovechar el uso de las unidades de procesamiento gráfico para acelerar el rendimiento del algoritmo que se implemente. Cabe destacar que los equipos presentes en el laboratorio del *GIDM* tienen disponible tarjetas gráficas con este soporte.

---

<sup>2</sup><http://www.intel.com>

<sup>3</sup><http://www.nvidia.com>

## Capítulo 3

# Detección de puntos característicos

### 3.1. Introducción

En el capítulo anterior se valoraron los distintos tipos de técnicas para el registro de imágenes, y en base a estos se planteó el uso de características locales para establecer la correspondencia entre imágenes. En el presente capítulo se pretende explicar en detalle el funcionamiento de los algoritmos de detección más usados en estas aplicaciones, así como también la importancia de los avances que cada uno introduce. Después se presentan los algoritmos de emparejamiento, así como también una explicación de su funcionamiento. Finalmente, con el objetivo de caracterizar cada uno de los detectores y emparejadores, se implementa un módulo para comparar el rendimiento de estos bajo distintas condiciones. Esto nos permitirá luego seleccionar el mejor algoritmo para cada aplicación.

## 3.2. Revisión teórica

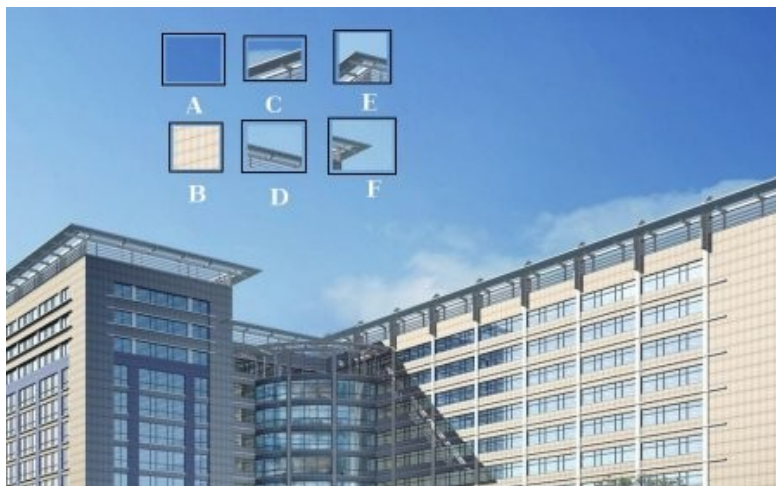
### 3.2.1. Detectores y descriptores de características

Antes mencionar la evolución de los algoritmos de detección de puntos de interés, primero es necesario definir que son estos. Los puntos de interés, puntos clave, o *"features"* (en español: características) como son comúnmente llamados, son regiones en una imagen que contienen patrones específicos, lo que hace que puedan ser fácilmente seguidos o ubicados en otra imagen. Tuytelaars y Mikolajczyk [26] definen un punto característico local como *"un patrón en la imagen que difiere de su vecindario directo"*. De esta forma, se considera que los puntos característicos deben proporcionar la posibilidad de ser identificados en diferentes imágenes con el objetivo de emparejarlos.

Para alcanzar este objetivo los detectores y extractores de puntos característicos deben cumplir con ciertas propiedades que les permita funcionar bajo distintas condiciones, en concreto se busca que estos algoritmos cumplan con las siguientes propiedades:

- **Robustez:** El algoritmo debe ser capaz de detectar la misma ubicación del punto característico independientemente ante cambios en la escala, rotación, traslación, iluminación, transformaciones geométricas, artefactos de compresión y ruido.
- **Repetibilidad:** El algoritmo debe ser capaz de detectar el mismo punto característico de la misma escena bajo cambios en el punto de vista.
- **Exactitud:** El detector debe localizar el punto característico de manera precisa (misma ubicación de píxel). Especialmente para tareas de alineación de imágenes.
- **Generalidad:** El algoritmo debe ser capaz de detectar puntos que pueden ser usadas en distintas aplicaciones, es decir, que detecte varios tipos de características (esquinas, burbujas, etc.)

- **Eficiencia:** El algoritmo debe ser capaz de detectar puntos característicos en nuevas imágenes a gran velocidad, para soportar aplicaciones en tiempo real.
- **Cantidad:** El algoritmo debe detectar todos, o casi todos los puntos característicos presentes en la imagen.

FIGURA 3.1: Caracterización de regiones en una imagen<sup>1</sup>

Llegados a este punto es necesario definir el funcionamiento de los algoritmos detectores, y que consideren estos como puntos característicos, basados en la definición previamente planteada. Atendiendo a la imagen 3.1, se puede observar que se caracterizan seis áreas de interés. Analizando estos segmentos, vemos que **A** y **B** corresponden con superficies planas, lo que hace que sea muy difícil identificar la ubicación exacta de estas superficies en la imagen original. Por otro lado, tenemos las regiones **C** y **D**, las cuales corresponden con bordes en la imagen, si bien, se puede limitar en gran medida el área de búsqueda hacia toda las regiones del mismo bordes, sigue siendo difícil acertar con la ubicación correcta. Por ultimo, analizando las regiones **E** y **F** tenemos que corresponden a esquinas de la imagen original, en este caso se puede identificar fácilmente la ubicación exacta de la región en la imagen.

A partir de esta idea, en la cual se consideran las esquinas como regiones fácilmente identificables en una imagen, en 1988 nace el primer algoritmo de detección

<sup>1</sup>[https://docs.opencv.org/3.0-beta/doc/py\\_tutorials/py\\_feature2d/py\\_features\\_meaning/py\\_features\\_meaning.html](https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_feature2d/py_features_meaning/py_features_meaning.html)

de puntos de interés llamado Detector de esquinas de Harris [8] (nombre original en inglés: Harris Corner Detector), y como su nombre lo indica está basado en la detección de esquinas.

Retomando el concepto planteado previamente, este detector busca la diferencia de intensidad de una región con su entorno directo, es decir, se detectará una esquina para aquellas regiones que presenten una alta variación de intensidad, al desplazar la ventana estudiada en cualquier dirección. En la figura 3.2 se puede apreciar visualmente como funciona esta ventana de búsqueda.

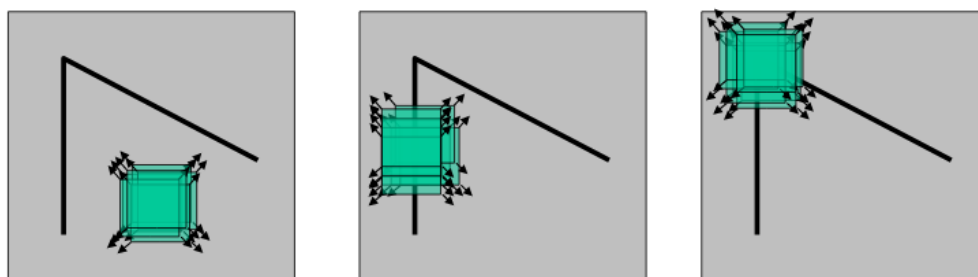


FIGURA 3.2: Ventana de búsqueda para de detección de esquinas, adaptado de<sup>2</sup>

Cuando se trabajan con detectores de características, se desea que estos sean invariantes ante la mayor cantidad de variables posibles, tal y como se menciona en la propiedad de robustez que deben tener estos algoritmos. Si bien el detector presentado anteriormente es invariante ante la traslación y la rotación (ya que las esquinas se mantienen como esquinas si son rotadas o desplazadas), no funciona de la misma forma ante cambios de escala. Como se observa en la figura 3.3, una región considerada como esquina, se podría considerar plana si es ampliada.

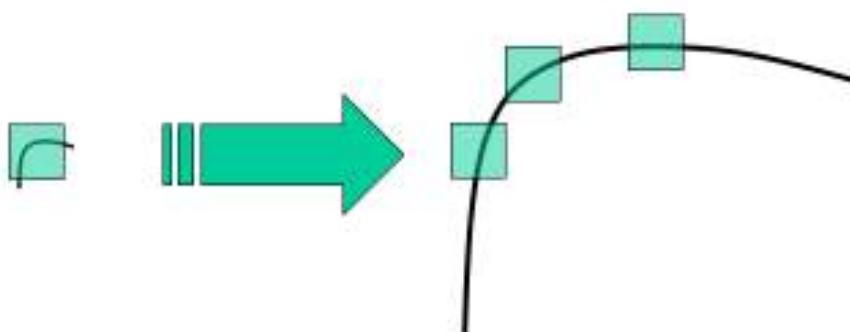


FIGURA 3.3: Efecto del escalado sobre las esquinas<sup>3</sup>

<sup>2</sup><https://dsp.stackexchange.com/questions/14338/>



Con el fin de conseguir detectar los mismos puntos ante cambios en la escala de la imagen, Lindeberg, T. [27] propone un algoritmo detector de “manchas” multi-escala a través de la búsqueda de un máximos en el espacio de escala, el cual se crea utilizando un operador laplaciano. El Laplaciano de Gaussianas *LoG* (del inglés: Laplacian-of-Gaussian), es una combinación lineal de segundas derivadas utilizado para detectar burbujas o manchas en una imagen. El funcionamiento es el siguiente: Dada una imagen de entrada, la representación para cada escala  $-s-$  de la imagen se define como la convolución de la imagen con un filtro Gaussiano con desviación estándar  $s$ .

Este resultado brinda una fuerte respuesta positiva para burbujas oscuras y respuestas fuertes negativas para burbujas claras, ambas de un tamaño  $2s$ , donde  $s$  es la escala. De esta forma las características detectadas presentan una fuerte relación entre el tamaño de las estructuras en la imagen y el grado de difusión del filtro gaussiano. Donde la desviación estándar del filtro se usa para controlar la escala cambiando que tanto se difumina la imagen.

Llegados aquí, una vez se haya detectado la ubicación de los puntos característicos en la imagen, la información de la localidad de este debe ser codificada y almacenada, y de esta forma lograr tener un descriptor único de la región con el objetivo final de ubicarlo en otra imagen. Con este fin se desarrollaron los algoritmos descriptores, los cuales una vez tengan la ubicación de los puntos característicos se encargan de convertir la información de su alrededor en una serie de números, o un vector que permita diferenciar un punto clave de otro. Esta información también es necesaria que sea invariante ante las variable mencionadas previamente, para lograr una identificación eficiente del mismo punto en distintas imágenes bajo distintas condiciones.

Partiendo de estos problemas, y del hecho que el calculo del operador LoG es computacionalmente costoso, en 2004 D. Lowe crea el detector y descriptor *SIFT* [9] (del inglés: Scale Invariant Feature Transform), en el cual el espacio de escala es construido en forma piramidal con la diferencia de gaussianas DoG (del inglés: Difference of Gaussians). En este sentido, El operador DoG ofrece una aproximación al LoG, donde se calcula sin convolución restando niveles de escala

---

<sup>3</sup>[https://docs.opencv.org/3.0-beta/doc/py\\_tutorials/py\\_feature2d/py\\_sift\\_intro/py\\_sift\\_intro.html](https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_feature2d/py_sift_intro/py_sift_intro.html)

adyacentes de una pirámide gaussiana. El proceso para la detección y descripción de puntos de interés de este algoritmo, consta de cuatro pasos principales:

En primer lugar, realiza una detección de máximos en el espacio de la escala aplicando la diferencia gaussiana *DoG*. Para esto, se aplica el filtro gaussiano con distintos tamaños de media (se tienen distintas escalas), luego restando estas imágenes para distintos pares de escalas se logra la diferencia de gaussiana. Posteriormente se buscan los máximos locales a lo largo del espacio (coordenadas X,Y) para cada correspondiente escala. Este proceso de detección se puede visualizar en la figura 3.4.

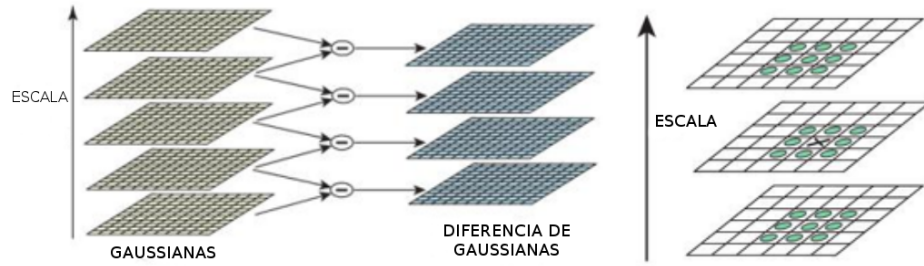


FIGURA 3.4: Detección de máximos en el espacio de la escala DoG, adaptado de [9]

En segundo lugar para la localización de puntos de interés, se descartan los puntos encontrados en el paso anterior que no superen cierto valor de umbral, es decir, que no estén lo suficientemente contrastados con su entorno. Con esta etapa el algoritmo solo toma en cuenta los puntos claves mas fuertes por cada escala. Además, con el objetivo de eliminar los bordes suficientemente contrastados que no correspondan con esquinas, el algoritmo usa una matriz hessiana para calcular las curvaturas principales, y así quedarse solo con esquinas.

Para garantizar la invarianza con respecto a la rotación, se toman los píxeles vecinos al punto clave y se calcula la magnitud y dirección del gradiente en esa región. Con esto se hace un histograma de la magnitud del gradiente en cada dirección, donde el pico mayor del histograma indica la orientación. En el caso que exista un pico mayor al 80 % del pico principal, este se utiliza para crear otro punto de interés en la misma posición pero con la distinta rotación.

Finalmente para crear el vector descriptor por cada punto clave se crea una matriz de 16x16 alrededor de este, dividida en 4 subregiones de 4x4 píxeles con un

histograma de orientaciones para cada uno. Seguidamente, el descriptor del punto será el vector con los valores de los histogramas de las regiones 4x4 concatenados. La figura 3.5 la representación del descriptor de SIFT.

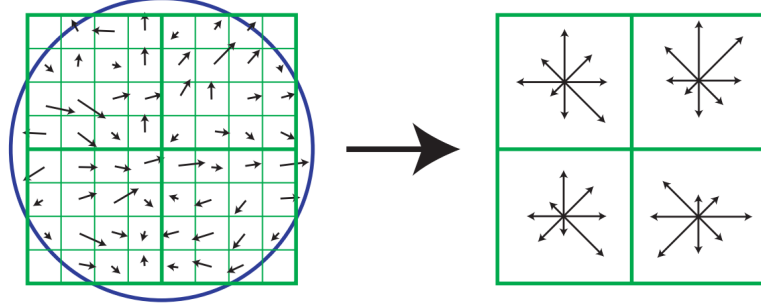


FIGURA 3.5: Izquierda: imagen de gradientes. Derecha: descriptor del punto clave. de [9]

En el año 2006, un grupo de tres personas Bay, H., Tuytelaars, T. and Van Gool, L. desarrollan *SURF* [10], el cual es un detector y descriptor de características basado en SIFT, pero con modificaciones que aumentan su velocidad de detección. Si bien, sacrifica un poco de rendimiento y precisión, lo hace mas provechoso para aplicaciones embebidas que demanden mayor velocidad de computo y menor uso de recursos, como por ejemplo *SLAM*. El proceso para la extracción de características por parte de este algoritmo se compone de los siguientes pasos:

Como primer paso, en lugar de aproximar el laplaciano de Gauss *LoG* (del inglés: Laplacian of Gaussians) con la diferencia de Gaussianas (DoG) como lo hace SIFT, este algoritmo aproxima LoG con cuadrados para promediar la imagen. La ventaja de aplicar filtros con cuadrados es que con la ayuda de imágenes integrales el cálculo computacional se reduce en gran medida.

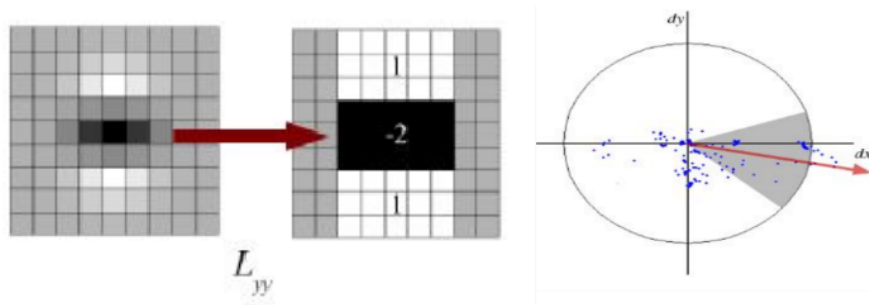


FIGURA 3.6: Izquierda: aproximación a la derivada de segundo orden del filtro gaussiano (derivada parcial en el eje  $y$ ) y su aproximación con un filtro cuadrado. Derecha: vector de orientación del descriptor. Adaptado de [10]

En función de identificar la orientación, el algoritmo utiliza la respuesta wavelet Haar en horizontal y vertical en un vecindario de  $6s$  (donde  $s$  es la escala evaluada) píxeles al rededor del punto de interés, Luego estas respuestas son representadas como puntos en el espacio, para luego calcular la orientación dominante con la suma de todos los resultados dentro de una ventana deslizante de apertura  $60^\circ$ . En la figura 3.6 se puede visualizar en el lado izquierdo, la aproximación que realiza de la derivada de segundo orden del filtro gaussiano, y su aproximación con un filtro cuadrado. Del lado derecho se ilustra el vector de orientación en función a la distribución de puntos estudiados.

El siguiente avance importante en los algoritmos de detección aparece en el año 2011 con *ORB* [13] (del inglés: Oriented FAST and Rotated BRIEF), este utiliza una combinación del detector FAST (del inglés: Features from Accelerated Segment Test) y del descriptor BRIEF (del inglés: Binary Robust Independent Elementary Features), este nuevo algoritmo esta caracterizado por su alta velocidad de procesamiento manteniendo un buen rendimiento, gracias al uso de un descriptor binario.

Como se mencionó utiliza el algoritmo FAST, el cual consiste en encontrar esquinas evaluando los píxeles en un perímetro circular, de esta forma, un punto será detectado como esquina si la cantidad de píxeles de color opuesto al evaluado, supera cierto valor de umbral (ver izquierda en la figura 3.7), posteriormente con el fin de aumentar la robustez, es aplicado el algoritmo de clasificación de esquinas de *Harris*. De igual forma se realiza con una estructura piramidal evaluando varias escalas (al igual que SIFT).

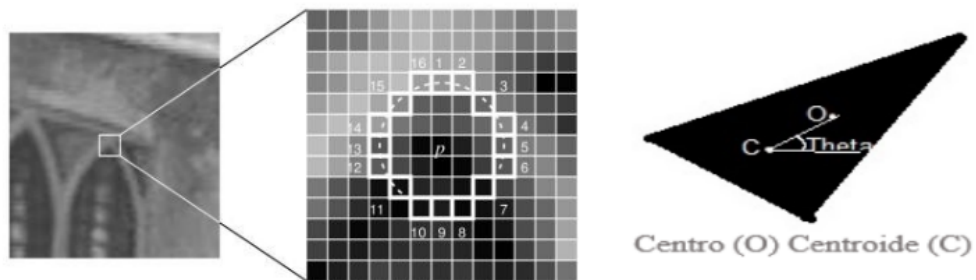


FIGURA 3.7: izquierda: detección de esquinas usando FAST, de [11]. derecha: descriptor basado en BRIEF, adaptado de<sup>4</sup>

<sup>4</sup><https://gilscvblog.com/2013/10/04/>

Como el algoritmo FAST no toma en cuenta la orientación, en el ORB se modificó para que calculara la orientación de la siguiente forma: Se considera una región ubicada en el centro del punto estudiado, luego se calcula el centroide de la región en función a la intensidad de los puntos. De esta forma, la dirección del vector desde el punto central hasta el centroide es asignado como vector de orientación. Observando a la derecha en 3.7 se aprecia un ejemplo del lugar del centroide ( $C$ ) y del centro ( $O$ ) para una región en particular.

Para el descriptor utiliza BRIEF, a diferencia de los anteriores (SIFT y SURF) este es un descriptor binario y no vectorial. El descriptor BRIEF produce una palabra de  $n$ -bits usando el algoritmo *Local Binary Tests* (LBT), el problema de esta representación es que no es muy robusta ante cambios en la rotación. Para resolver esto ORB utiliza la información de la orientación previamente calculada en el paso de detección para aplicar LBT en esa orientación.

Los algoritmos de detección que se mencionaron hasta este momento tienen una característica en común, y es que cuando trabajan con el esquema piramidal lo hacen bajo el espacio de escala Gaussiano, el cual es una instancia particular de difusión lineal. De esta forma, al utilizar este filtro no se respetan los límites naturales de los objetos y se difumina del mismo nivel toda la región de la imagen cuando se avanza entre niveles de escala.

Enfocándose en esta característica, en el año de 2012 se desarrolla el detector y descriptor llamado KAZE [14] por parte de *Pablo Fernández Alcantarilla*. Este novedoso algoritmo opera completamente en un espacio de escala no lineal, y para ello utilizan un esquema de división de operadores aditivos (AOS, del inglés: Additive Operator Splitting), que les permite obtener espacios de escala no lineales de forma eficiente. De este modo se puede realizar un difuminado localmente adaptativo, posibilitando que se remueva el ruido en las imágenes, manteniendo información importante sobre los bordes de los objetos al avanzar en el espacio de escala. En la figura 3.8 se puede observar como afecta en los bordes de los objetos el aplicar un filtro de difusión lineal, y uno que no lo es, bajo el esquema propuesto por este algoritmo.

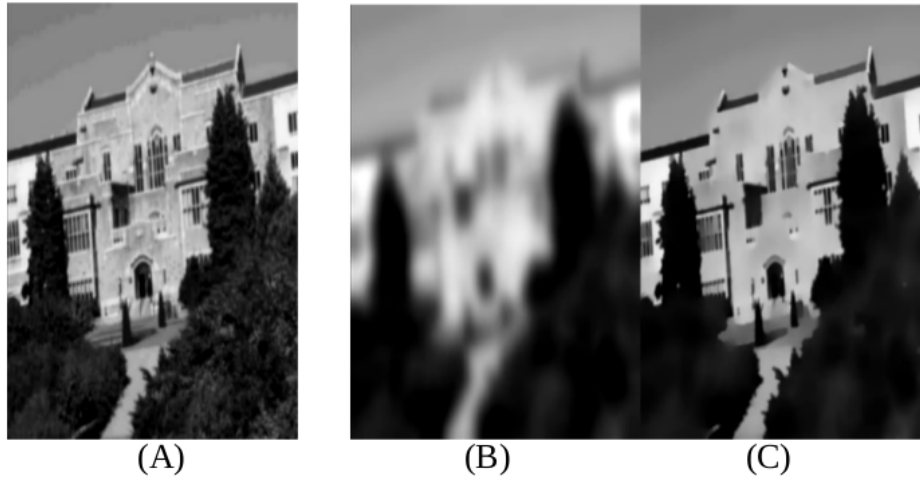


FIGURA 3.8: (A): imagen original, (B) filtro lineal Gaussiano, (C) filtro no lineal usado en KAZE, adaptada de [14]

Bajo este mismo esquema de difusión no lineal, el mismo autor en el año 2013 desarrolla la versión acelerada de este algoritmo que recibe el nombre de *A-KAZE* [15] (del inglés: Accelerated KAZE). Esta mejora se utiliza un esquema basado en difusión explícita rápida *FED* (del inglés: Fast Explicit Diffusion) en lugar de *AOS*, el cual es un nuevo esquema piramidal que incrementa en gran medida la velocidad de cómputo para construir el espacio de escala no lineal.

Para el cálculo de la orientación el primer algoritmo *KAZE* utiliza un descriptor para la orientación similar al que emplea SURF. Este encuentra la orientación dominante en un área circular de radio  $6s$  ( $s$  corresponde con la escala), y para cada muestra del círculo se calcula la derivada de primer orden en las direcciones  $X$  e  $Y$ , y se ponderan con una gaussiana centrada en el punto de interés. Luego, las respuestas de estas derivadas son representadas como puntos en un espacio vectorial, donde la orientación dominante se haya sumando las respuestas dentro de un segmento de círculo deslizante con apertura de  $60^\circ$ .

Por otro lado, la versión acelerada *A-KAZE* emplea un descriptor basado en una versión modificada del algoritmo de diferencia local binaria *LDB* [28] (del inglés: Local Difference Binary), llamado M-LBD (del inglés: Modified Local Difference Binary), el cual aprovecha al máximo la información del espacio de escala no lineal. La modificación consiste en hacer un sub-muestreo de cada región que divide la zona del descriptor, en lugar de calcular el promedio de todos los píxeles de la región, es decir, se tienen muestras de cada subdivisión para distintas escalas.

### 3.2.2. Emparejadores de puntos característicos

En este punto ya hemos estudiado los distintos algoritmos que permiten encontrar y clasificar puntos de interés en las imágenes. Ahora bien, es necesario identificar cuales de estos puntos corresponden con la misma ubicación, tal y como podemos observar en la figura 3.9. Para establecer esta relación se utilizan los algoritmos emparejadores de características, los cuales relacionan estos puntos en base a los descriptores de dichos puntos.

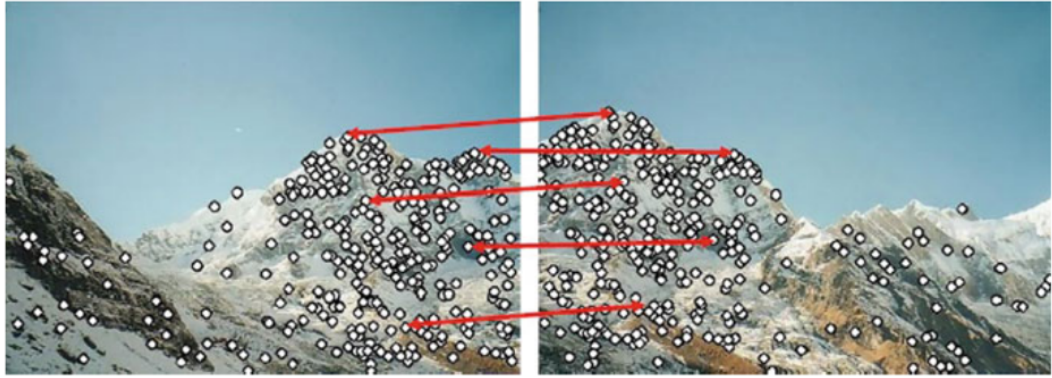


FIGURA 3.9: Emparejamiento de puntos que corresponden a la misma ubicación, de [29]

El proceso para realizar el emparejamiento consiste en el siguiente: Teniendo un punto característico  $P_1$  perteneciente a la imagen 1, y por otro lado se teniendo un punto  $P_2$  perteneciente a la imagen 2, Se calcula la distancia entre los descriptores  $D_1$  y  $D_2$ . En el caso de descriptores vectoriales, esta distancia corresponde con la distancia *Euclidiana*, dada por la siguiente expresión:

$$D = \sqrt{(v_1 - q_1)^2 + (v_2 - q_2)^2 + \dots + (v_n - q_n)^2}$$

Donde  $v_n$  corresponden con los componentes del vector descriptor  $D_1$ ,  $q_n$  con los componentes del vector descriptor  $D_2$ , y  $n$  en el numero de componentes de ambos vectores. Por otro lado, para los descriptores binarios, se calcula la distancia *Hamming*, dada por la siguiente expresión:

$$D = ||D_1 \oplus D_2||$$

Este proceso se repite hasta tener la distancia de cada punto de la imagen 1 con todos los puntos de la imagen 2, y viceversa. Al tener estas distancias, los puntos se emparejarán si y solo si se cumplen las siguientes condiciones:

- (I) El punto  $P_1$  presenta la mejor distancia con  $P_2$ , en relación a todos los puntos de la imagen 2.
- (II) El punto  $P_2$  presenta la mejor distancia con  $P_1$  en relación a todos los puntos de la imagen 1.

Este proceso es mejor conocido como emparejamiento por fuerza bruta, ya que se compara entre todos los puntos por la mejor pareja posible. Si bien se asegura obtener el mejor emparejamiento, siendo viable para trabajar con pocos datos, el hecho de probar todas los casos posibles cuando se tiene una gran cantidad de puntos, incrementa en gran medida el tiempo de computo. Para efectuar este proceso de una forma eficiente se desarrollaron algoritmos basados en la búsqueda de vecinos mas cercanos. En este sentido se cuenta con el algoritmo *kd-forest* (abreviado del inglés: k-dimensional forest), el cual es una mejora del algoritmo *kd-tree* (abreviado del inglés: k-dimensional tree) para mejor desempeño al usar vectores multidimensionales, implementado en la librería para la rápida aproximación de vecinos mas cercanos FLANN [30] (del inglés: Fast Library for Approximate Nearest Neighbors).

### 3.3. Módulo comparativo

#### 3.3.1. Resultados

### 3.4. Conclusiones

Resumen



# Capítulo 4

## Alineación de imágenes

### 4.1. Introducción

Introducción

### 4.2. Revisión teórica

Prueba

### 4.2.1. Transformaciones geométricas

## 4.3. Generación de sub-mosaicos

---

### Algoritmo 1: Registro de imágenes

---

```

1  $I_{i+1} \equiv$  Imagen nueva;
2  $I_i \equiv$  ultima imagen anadida al mosaico;
3  $V_i \equiv$  vecinos de  $I_i$ ;
4 mientras puntos emparejados  $\geq 4$  hacer
5     Emparejar puntos de  $I_{i+1}$  con  $I_i$ ;
6     si  $I_i$  tiene vecinos entonces
7         para cada vecino de  $I_i$  hacer
8             Emparejar puntos de  $I_{i+1}$  con  $V_i$ ;
9         fin
10    fin
11    descartar malos emparejamientos;
12    aplicar busqueda sectorizada;
13    si puntos totales emparejados  $\leq 3$  entonces
14        modificar criterio para descartar;
15        si criterio para descartar llega al minimo entonces
16            terminar ;                               // no es posible emparejar imagen
17        fin
18    fin
19 fin

```

---

### 4.3.1. Selección de imagen de referencia

Prueba

### 4.3.2. Matriz de transformación promedio

---

**Algoritmo 2:** Calculo de matriz de homografía promedio

---

```
20 mientras no se alcanza el maximo de iteraciones hacer
21     seleccionar 4 puntos aleatorios del primer sub-mosaico;
22     seleccionar los 4 puntos correspondientes en el segundo sub-mosaico;
23     calcular el punto medio para cada par de puntos correspondientes;
24     calcular la transformacion desde los puntos del primer sub-mosaico hasta los
        puntos medios;
25     aplicar transformacion en el primer sub-mosaico;
26     calcular error de distorsión en el primer sub-mosaico;
27     si el error es menor que el mas bajo obtenido entonces
28         guardar el error como el mas bajo;
29         guardar la matriz de transformación como la mejor;
30     en otro caso
31         restaurar valores del primer sub-mosaico;
32     fin
33 fin
```

---

## 4.4. Corrección euclidiana

Prueba

## 4.5. Resultados

Resumen

## 4.6. Conclusiones

Resumen

# Capítulo 5

## Unión de imágenes

### 5.1. Introducción

Prueba

### 5.2. Linea de costura

Prueba

#### 5.2.1. Mapa mas cercano

Prueba

#### 5.2.2. Corte por grafo

Prueba

## **5.3. Corrección de color**

Prueba

### **5.3.1. Ajuste de ganancia**

Prueba

### **5.3.2. Método de Reinhard**

Prueba

## **5.4. Fusión de imágenes**

### **5.4.1. Fusión ponderada**

Prueba

### **5.4.2. Fusión piramidal**

Prueba

## **5.5. Resultados**

## **5.6. Conclusiones**

## Capítulo 6

# Conclusiones y trabajos futuros

Conclusiones

# Bibliografía

- [1] Danilo Díaz Tarascó. Desarrollo de sistema de telemetria para teleoperación de robot submarino, 2016.
- [2] Said Alexander Alvarado Marín. Caracterización, instrumentación y control de robot cuadricóptero volador, 2017.
- [3] Paul R. Wolf. *Elements of Photogrammetry (Second Edition)*. McGraw-Hill Higher Education, 1983.
- [4] D.P. Capel. *Image Mosaicing and Superresolution*. Springer Science and Business Media, 2004.
- [5] Sherin Ghannam and A. Lynn Abbott. Cross correlation versus mutual information for image mosaicing. *International Journal of Advanced Computer Science and Applications*, 2013.
- [6] Kostas Berberidis y Irene Karybali. A new efficient cross-correlation based image registration technique with improved performance. *European Signal Processing Conference*, 2002.
- [7] Amaury Dame y Eric Marchand. Video mosaicing using a mutual information-based motion estimation proces. *IEEE International Conference on Image Processing (ICIP)*, 2011.
- [8] Hans P. Morevec. Towards automatic visual obstacle avoidance. 1977.
- [9] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Computer Science Department, University of British Columbia*, 2004.
- [10] Tuytelaars T. y Van Gool L Bay, H. Surf: Speeded up robust features. *European Conference on Computer Vision*, 2006.

- [11] Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. *Proceedings of the 9th European Conference on Computer Vision - Volume Part I*, 2006.
- [12] Vincent y Strecha Christoph y Fua Pascal Calonder, Michael y Lepetit. Brief: Binary robust independent elementary features. In *Proceedings of the 11th European Conference on Computer Vision: Part IV*. Springer-Verlag, 2010.
- [13] Kurt Konolige y Gary Bradski Ethan Rublee, Vincent Rabaud. Orb: an efficient alternative to sift or surf. *International Conference on Computer Vision*, 2011.
- [14] y Andrew J. Davison Pablo Fernandez Alcantarilla, Adrien Bartoli. Kaze features. *European Conference on Computer Vision*, 2012.
- [15] Pablo Fernández Alcantarilla. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *British Machine Vision Conference (BMVC)*, 2013.
- [16] G. Wiet T. Pan y K. Huang J. Prescott, M. Clary. Automatic registration of large set of microscopic images using high-level features. *IEEE International Symposium on Biomedical Imaging*, 2006.
- [17] G.R. Keller H. Huang y V. Kreinovich H. Xie, N. Hicks. An idl/envi implementation of the fft-based algorithm for automatic image registration. *Computers and Geosciences*, 2003.
- [18] Z. Zhang y H. Tang F. Yang, L. Wei. Image mosaic based on phase correlation and harris operator. *Journal of Computational Information Systems*, 2012.
- [19] U.C. Pati A. Pandey. A novel technique for non-overlapping image mosaicing based on pyramid method. *IEEE India Conference*, 2013.
- [20] Fan Zhang and Feng Liu. Parallax-tolerant image stitching. *IEEE Conference on Computer Vision and Pattern Recognition*, 2014.
- [21] Shahriar Negahdaripour y Mohammad Mahoor Nuno Gracias, Art Gleason. Fast image blending using watersheds and graph cuts. *Image and Vision Computing* 27(5):597-607, 2009.
- [22] Y. Boykov y V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001.



- 
- [23] Fabio Pazzaglia y Carlo Colombo Fabio Bellavia, Marco Fanfani. Piecewise planar underwater mosaicing. *OCEANS'15 MTS/IEEE Genova*, 2015.
  - [24] Fabio Bellavia and Carlo Colombo. Estimating the best reference homography for planar mosaics from videos. *VISAPP - 10th International Conference on Computer Vision Theory and Applications; VISIGRAPP, Proceedings*, 2015.
  - [25] Armagan Elibol, Jinwhan Kim, Nuno Gracias, and Rafael García. Fast underwater image mosaicing through submapping. *Journal of Intelligent and Robotic Systems*, 85:167–187, 2017.
  - [26] Tinne Tuytelaars y Krystian Mikolajczyk. Local invariant feature detectors: a survey. *Foundations and Trends® in Computer Graphics and Vision*, 2008.
  - [27] T Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 1998.
  - [28] Kwang-Ting Cheng Xin Yang. Local difference binary for ultrafast and distinctive feature description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.
  - [29] Richard Szeliski. *Computer Vision: Algorithms and Applications*. Springer-Verlag New York, Inc., New York, NY, USA, 1st edition, 2010.
  - [30] David Muja, Marius y Lowe. Scalable nearest neighbor algorithms for high dimensional data. 36:2227–2240, 11 2014.

# Apéndice A

## Instalación de librería OpenCV y dependencias

### A.1. @sección

#### A.1.1. @subsección

“Saludo”.