

UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR



DETECCIÓN DE LA POSICIÓN RELATIVA DE UNA CÁMARA MÓVIL EN SITUACIONES DE GRANDES DESPLAZAMIENTOS.

María Narvárez Encinal.
Tutor: Marcos Escudero Viñolo
Ponente: Jesús Bescós Cano

-TRABAJO FIN DE GRADO-

Departamento de Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Junio 2015

DETECCIÓN DE LA POSICIÓN RELATIVA DE UNA CÁMARA MÓVIL EN SITUACIONES DE GRANDES DESPLAZAMIENTOS.

María Narváez Encinal

Tutor: Marcos Escudero Viñolo

Ponente: Jesús Bescós Cano



Video Processing and Understanding Lab

Departamento de Ingeniería Informática

Escuela Politécnica Superior

Universidad Autónoma de Madrid

Junio 2015

Resumen

Este proyecto tiene como objetivo final la implementación de algoritmos que sean capaces de modelar, mediante estimaciones de la matriz fundamental, la geometría epipolar existente entre dos imágenes. La finalidad de este proceso es la obtención de manera automática de la posición de las cámaras mediante los parámetros extrínsecos, es decir, la calibración externa automática de dichas cámaras.

Para la consecución del objetivo principal se ha estudiado previamente el comportamiento de las cámaras a la hora de capturar y formar imágenes así como diferentes métodos de detección, descripción de puntos y búsqueda de correspondencias entre imágenes. Entre ellos, podemos destacar el detector D_{1norm} que mide la característica Hessiana sobre el Scale-Space, y los algoritmos SIFT-DSIFT y DAISY como descriptores y NRDC para el establecimiento de correspondencias densas.

Ayudándonos de los métodos anteriores se obtiene una primera estimación de la matriz fundamental mediante el algoritmo RANSAC. Posteriormente, se realizará un refinamiento de dicha estimación inicial a partir de un conjunto de técnicas diseñadas y desarrolladas en el transcurso del trabajo fin de grado que se presenta.

Adicionalmente, se ha diseñado una herramienta que permite al usuario seleccionar manualmente las correspondencias utilizadas para el cálculo de la primera estimación de la matriz fundamental y, que a partir de dicha estimación, pueda desarrollar la fase de refinamiento. Esta herramienta será de utilidad cuando la extracción automática de correspondencias no sea posible debido a las condiciones de captura de la escena.

Por último, se han evaluado las estimaciones iniciales de la matriz fundamental y las obtenidas tras la fase de refinamiento a través de diferentes medidas de error.

Palabras clave

Matriz fundamental, geometría epipolar, detección de puntos, descripción, búsqueda de correspondencias, matriz de proyección

Abstract

This project aims to design and develop a set of algorithms able to cope with the problem of modelling the epipolar geometry between two cameras capturing the same scene. The ultimate goal of this process is the automatic estimation of the relative position between both cameras. Whereas the epipolar geometry is fully explained through the fundamental matrix, the relative position may be extracted, up to a scene homography, from the fundamental matrix.

To fulfill the faced task, an study of the image capture and camera calibration processes was first required. Automatic calibration strategies usually rely in a previous set of inter-image correspondences, hence, the analysis of this research topic was also required during this project. Specifically, Hessian-based detectors on the Scale-Space were explored for the feature detection stage, SIFT-DSIFT and DAISY descriptors were assessed to describe the extracted features, whereas the NRDC algorithm was evaluated as an alternative for feature matching.

The use of these feature-matching pipeline leads—in conjunction with the RANSAC estimator—to an initial estimation of the fundamental matrix. Afterwards, such matrix is refined by a set of strategies designed and developed during this project.

Moreover, a Graphical User Interface (GUI) which eases the manually introduction of reliable correspondences has been also designed and developed. The aim of this GUI is to face scenarios where the feature-matching pipeline is not able to produce an initial estimation of enough quality,

Finally, both the initial estimates of the fundamental matrix and its subsequent refinements have been evaluated by means of different experiments and error measurements.

Key words

Fundamental matrix, epipolar geometry, feature detection and description, correspondence searching, projection matrix.

Agradecimientos

En primer lugar, y como no podía ser de otra manera, me gustaría dar las gracias de corazón a mi tutor, Marcos, por su tiempo, su dedicación y, simplemente, por darme la oportunidad de trabajar con él. Ha sido todo un placer.

No puedo olvidarme de Chema y Jesús, cuya labor como profesores despertó en mí el interés por esta rama de las telecomunicaciones. Gracias por permitirme trabajar en este laboratorio.

De igual manera, quiero agradecer al resto de compañeros del VPU, por hacer más amenos los días de trabajo y por su entera disposición para cualquier cosa.

Por supuesto, gracias a todos y cada uno de mis compañeros durante estos años, por compartir agobios, apuntes y conocimientos. En especial a ellas, las que me han hecho vivir los mejores momentos de esta etapa. Por un lado Nerea, mi pelirroja, y nuestras ansias constantes de cespéd. Qué tiempos, cualquier excusa era buena para no hacer nada. Por otro lado mi brasileña, cuyo ánimo diario ha ayudado a que hoy termine mi TFG. Por saber estar y saber escuchar, muchísimas gracias Marta. Por último, y pese a todo, está mi gran debilidad. Sandra, gracias por hacer de las innumerables horas de prácticas algo llevadero, por los descansos, por las clases y por absolutamente todo. Ojalá se pudiera retroceder en el tiempo.

También quiero hacer especial mención a los de siempre. Mis Caurienses, por hacer de cada momento libre el mejor y por estar siempre a mi lado a pesar de los 300 km que nos separan.

Lo mejor siempre se deja para el final así que, por último, gracias a mi familia. A mis abuelos, simplemente porque deberían ser eternos. Pero en especial a mis padres y a mi hermana, porque son ellos los que realmente han hecho esto posible; por su apoyo incondicional, por confiar ciegamente en mí, por sus abrazos y sus “Venga que ya no te queda nada”, en definitiva, por hacer de mí la persona que soy ahora. Gracias, no sabéis cuánto os quiero.

María Narváez Encinal. Junio 2015.

Índice general

Resumen	v
Abstract	vii
Agradecimientos	ix
1. Introducción.	1
1.1. Motivación.	1
1.2. Objetivos.	2
1.3. Estructura de la memoria.	2
2. Estado del arte y conceptos básicos	5
2.1. Geometría Proyectiva	5
2.1.1. Modelo de cámara pin-hole	6
2.1.2. Matriz de Proyección	8
2.1.3. Geometría Epipolar y Matriz Fundamental	12
2.2. Detección de puntos de interés	17
2.2.1. Scale-Space	17
2.2.2. Detectores sobre el Scale-Space	18
2.3. Descripción de puntos de interés	19
2.3.1. SIFT (Scale-Invariant Feature Transform) y DSIFT (Dense Scale-Invariant Feature Transform)	20
2.3.2. DAISY	21
2.4. Correspondencias entre imágenes	22
2.4.1. Método de obtención de correspondencias	22
2.4.2. NRDC (Non-Rigid Dense Correspondence)	22
3. Diseño y desarrollo	25
3.1. Introducción	25
3.2. Búsqueda de correspondencias iniciales	25
3.2.1. Correspondencias locales	26
3.2.2. Correspondencias densas	28
3.3. Estimación de la matriz fundamental inicial	28
3.4. Refinamiento	28
3.4.1. Mejora de la descripción	28

3.4.2. Mejora de las correspondencias	29
3.5. Reestimación de la matriz fundamental	32
3.6. Rotación y Traslación	32
3.7. Limitaciones	33
4. Interfaz gráfica de usuario (GUI)	35
5. Resultados experimentales	39
5.1. Dataset y Código	39
5.2. Procedimiento de evaluación	40
5.3. Extracción de la información de <i>Ground-truth</i>	40
5.4. Medidas de error	42
5.4.1. Error de las correspondencias con Ground-Truth	42
5.4.2. Error de la matriz fundamental usando información de Ground-Truth	43
5.4.3. Error de las correspondencias sin utilizar información de Ground-Truth	43
5.5. Análisis de resultados	45
5.5.1. Análisis de F_0 a partir de la búsqueda de correspondencias iniciales	45
5.5.2. Análisis de F tras el refinamiento	48
5.6. Estimación de la posición de las cámaras	54
5.7. Evaluación de resultados	57
6. Conclusiones y trabajo futuro.	61
6.1. Conclusiones.	61
6.2. Trabajo futuro.	62
Bibliografía	64
A. Transformaciones para el cálculo de P	67
B. Imágenes evaluadas	71
C. Tablas de resultados	73
C.1. Estimación inicial SIFT	73
C.2. Refinamiento	75
C.2.1. Error de las matrices fundamentales	76
C.2.2. Errores de las correspondencias	77

Índice de figuras

2.1.	<i>Modelo de cámara pin-hole.</i>	7
2.2.	<i>Transformación Euclídea en 3D.</i>	10
2.3.	<i>Rotación de los ejes individualmente.</i>	10
2.4.	<i>Geometría epipolar</i>	12
2.5.	<i>Ejemplo de Scale-Space.</i>	18
2.6.	<i>Orientaciones del descriptor SIFT.</i>	20
2.7.	<i>Descriptor DAISY.</i>	21
2.8.	<i>Descripción del algoritmo NRDC.</i>	23
3.1.	<i>Diagrama de bloques.</i>	25
3.2.	<i>Ejemplo de detección de puntos sobre el Scale-Space con descripción SIFT.</i>	27
3.3.	<i>Proyección.</i>	30
3.4.	<i>Retroproyección.</i>	30
3.5.	<i>Búsqueda exhaustiva.</i>	31
3.6.	<i>Proyección y retroproyección con uso de máscara.</i>	32
4.1.	<i>Interfaz de usuario.</i>	36
4.2.	<i>Selección de imágenes, estimación inicial y refinamiento.</i>	37
4.3.	<i>Obtención de resultados</i>	38
5.1.	<i>Extracción de los mapas de error de F.</i>	41
5.2.	<i>Ejemplo de mapa de error de la matriz fundamental.</i>	44
5.3.	<i>Comparativa de los dos métodos de estimación inicial.</i>	48
5.4.	<i>Búsqueda de nuevas correspondencias con fase inicial estimada a partir de correspondencias locales.</i>	49
5.5.	<i>Búsqueda de nuevas correspondencias con fase inicial estimada a partir de correspondencias densas.</i>	49
5.6.	<i>Mapas de error utilizando búsqueda de correspondencias locales para la estimación inicial F_0.</i>	50
5.7.	<i>Mapas de error utilizando búsqueda de correspondencias densas para la estimación inicial F_0.</i>	51
5.8.	<i>Leyenda de color.</i>	51
5.9.	<i>Errores en las correspondencias analizando la técnica de Proyección y Retroproyección.</i>	52

5.10. Errores en las correspondencias analizando la técnica de <i>Búsqueda Exhaustiva</i>	52
5.11. Errores en las correspondencias analizando la técnica que combina los métodos anteriores.	52
5.12. Error de los vectores de traslación para una fase de estimación inicial mediante correspondencias densas.	55
5.13. Error de los vectores de traslación para una fase de estimación inicial mediante correspondencias locales.	55
5.14. Error de los vectores de rotación para una fase de estimación inicial mediante correspondencias densas.	56
5.15. Error de los vectores de rotación para una fase de estimación inicial mediante correspondencias locales.	56
A.1. Cambio del sistema de coordenadas de la imagen al sistema de coordenadas normalizado.	68
B.1. Imágenes a analizar.	71

Índice de tablas

5.1.	<i>Búsqueda de correspondencias locales: barrido de los parámetros de configuración.</i>	46
5.2.	<i>Resultados de F_0 a partir de búsqueda de correspondencias locales.</i>	47
5.3.	<i>Resultados de F_0 a partir de búsqueda de correspondencias densas.</i>	47
5.4.	<i>Estimación inicial de F mediante correspondencias manuales.</i>	53
5.5.	<i>Refinamiento de correspondencias manuales: Descriptor DSIFT.</i>	53
5.6.	<i>Refinamiento de correspondencias manuales: Descriptor DAISY.</i>	54
C.1.	<i>Búsqueda de correspondencias locales: barrido de los parámetros de configuración entre las imágenes 0000-0002.</i>	73
C.2.	<i>Búsqueda de correspondencias locales: barrido de los parámetros de configuración entre las imágenes 0000-0003.</i>	74
C.3.	<i>Búsqueda de correspondencias locales: barrido de los parámetros de configuración entre las imágenes 0000-0004.</i>	74
C.4.	<i>Búsqueda de correspondencias locales: barrido de los parámetros de configuración entre las imágenes 0000-0005.</i>	74
C.5.	<i>Búsqueda de correspondencias locales: barrido de los parámetros de configuración entre las imágenes 0000-0006.</i>	75
C.6.	<i>Búsqueda de correspondencias locales: barrido de los parámetros de configuración entre las imágenes 0000-0007.</i>	75
C.7.	<i>Resultados completos del refinamiento de la matriz fundamental extraída mediante correspondencias iniciales locales con el descriptor DAISY.</i>	76
C.8.	<i>Resultados completos del refinamiento de la matriz fundamental extraída mediante correspondencias iniciales locales con el descriptor DSIFT.</i>	76
C.9.	<i>Resultados completos del refinamiento de la matriz fundamental extraída mediante correspondencias iniciales densas con el descriptor DAISY.</i>	77
C.10.	<i>Resultados completos del refinamiento de la matriz fundamental extraída mediante correspondencias iniciales densas con el descriptor DSIFT.</i>	77
C.11.	<i>Correspondencias iniciales locales y fase de refinamiento DSIFT.</i>	77
C.12.	<i>Correspondencias iniciales locales y fase de refinamiento DAISY.</i>	78
C.13.	<i>Correspondencias iniciales densas y fase de refinamiento DSIFT.</i>	78
C.14.	<i>Correspondencias iniciales densas y fase de refinamiento DAISY.</i>	78
C.15.	<i>Errores en las correspondencias utilizando la técnica de Proyección y Retroproyección.</i>	78

C.16. <i>Errores en las correspondencias utilizando la técnica de Búsqueda Exhaustiva.</i>	78
C.17. <i>Errores en las correspondencias utilizando la técnica de Combinación de ambas.</i>	79

Capítulo 1

Introducción.

1.1. Motivación.

Actualmente los métodos más utilizados para la calibración de cámaras son aquellos que determinan sus parámetros a partir de puntos 3D conocidos y sus proyecciones en el plano de la imagen. Estas correspondencias son extraídas manualmente mediante el uso de patrones, centrandos su atención en la exactitud de los parámetros intrínsecos y extrínsecos obtenidos, pero olvidándose de la automatización del proceso.

Los parámetros intrínsecos definen la geometría y la óptica de las cámaras y los extrínsecos la posición y la orientación de éstas. Por lo tanto, la calibración es un procedimiento que intenta conocer cómo una cámara proyecta un punto 3D sobre el plano de la imagen en 2D y cómo se relacionan las proyecciones de los puntos 3D en varias cámaras.

Los parámetros extrínsecos puede obtenerse mediante el cálculo de la matriz fundamental, es decir, modelando la geometría epipolar existente entre dos (o más) imágenes capturadas por las cámaras (cada cámara una imagen). Una vez obtenida la matriz fundamental, es posible la extracción de las matrices de proyección de ambas cámaras (salvo como veremos por una homografía de escena) y, con ellas, los parámetros mencionados anteriormente estableciendo una de las cámaras como cámara de referencia.

En definitiva, conocer esta geometría epipolar permite establecer una restricción proyectiva fuerte entre las proyecciones de los objetos 3D en diferentes vistas. Esta restricción puede ser utilizada para mejorar el establecimiento de correspondencias biunívocas entre las proyecciones, que pueden ser usadas para, mediante triangulación, la obtención de la posición de objetos del mundo real.

1.2. Objetivos.

El objetivo principal de este trabajo es el diseño, implementación y refinamiento de técnicas que permitan estimar correctamente la matriz fundamental existente entre dos imágenes capturadas por cámaras ampliamente separadas. Esto implicaría la extracción de la geometría epipolar sin necesidad de conocer los parámetros intrínsecos y extrínsecos de las cámaras con las que se han capturado las imágenes y, además, permitiría la obtención de la posición de una de las cámaras con respecto de la otra.

Para la consecución del objetivo final, podemos desglosar los objetivos de la siguiente forma:

1. **Estudio detallado del estado del arte.** Este estudio puede dividirse en dos fases; en la primera de ellas, se estudiará y entenderá el proceso de captura y formación de imágenes por parte de las cámaras. En la segunda fase, se estudiarán las técnicas conocidas para la extracción de correspondencias entre imágenes.
2. **Implementación del código.** Se han desarrollado diferentes estrategias para la consecución del objetivo principal. Estas técnicas podrán combinarse para generar diferentes resultados.
3. **Creación de una interfaz de usuario.** Hemos generado una interfaz que permita al usuario generar las correspondencias que estimarán una primera aproximación de la matriz fundamental y, a partir de ella, puedan interactuar con las diferentes técnicas desarrolladas.
4. **Evaluación de resultados.** Evaluaremos las matrices fundamentales iniciales, así como las reestimaciones obtenidas mediante las técnicas generadas, y las compararemos entre sí.

1.3. Estructura de la memoria.

La memoria del trabajo se divide en los siguientes capítulos:

- Capítulo 1. Motivación, objetivos del proyecto y estructura de la memoria.
- Capítulo 2. Estudio de la geometría proyectiva. Estado del arte de las técnicas en detección y descripción de puntos de interés, así como de la obtención de las correspondencias entre imágenes.
- Capítulo 3. Diseño y desarrollo. Descripción de los métodos elegidos para la estimación inicial de la matriz fundamental. Explicación de las técnicas desarrolladas para el refinamiento.

- Capítulo 4. Herramienta para el refinamiento de correspondencias manuales. Creación de la interfaz usuario.
- Capítulo 5. Resultados experimentales. Análisis comparativo de todas las posibles combinaciones implementadas.
- Capítulo 6. Conclusiones derivadas del análisis de los resultados y enumeración de posibles líneas de trabajo futuro.
- Referencias y Apéndices.

Capítulo 2

Estado del arte y conceptos básicos

Este capítulo proporciona una visión general del trabajo realizado previamente en las áreas relacionadas con el objetivo del proyecto. Con las siguientes secciones entenderemos la estructura del trabajo así como los pasos a seguir para la aproximación y extracción de la matriz fundamental y, por lo tanto, de los parámetros extrínsecos necesarios para la estimación de la posición de una cámara respecto de la otra.

2.1. Geometría Projectiva

En este trabajo partimos de dos puntos de vista diferentes de una misma escena y, a partir de estas dos imágenes, tratamos de encontrar la posición de una cámara respecto de la otra. Para entender esto es necesario hacer una breve introducción a los fundamentos de captura y formación de imágenes ya que, a la hora de proyectar el mundo real sobre una imagen, perdemos una dimensión, pasando de las tres dimensiones (3D) del mundo que conocemos a las dos dimensiones (2D) que se representan en las imágenes. Para ello explicaremos el modelo de cámara pin-hole, los parámetros intrínsecos y extrínsecos de una cámara y la composición de la matriz de proyección. Una vez hecho esto, se tendrán los conocimientos necesarios para entender el concepto de geometría epipolar entre dos puntos de vista. Como veremos, esta geometría es completamente independiente de la estructura de la escena, dependiendo únicamente de los parámetros intrínsecos de la cámara y de la posición relativa de una cámara respecto de la otra. La denominada matriz fundamental contiene toda la información necesaria para expresar dicha geometría.

2.1.1. Modelo de cámara pin-hole

Los fundamentos de la formación de imágenes se basan generalmente en un modelo ideal de cámara denominado cámara pin-hole. El modelo de cámara pin-hole, cuya explicación podemos encontrar en varios de los artículos referenciados ([1], [2], [3]), consiste en una lente delgada cuya apertura tiende a cero, de esta forma, nos quedamos con el modelo más sencillo posible debido a que todos los rayos pasan por el centro óptico simplificando considerablemente el análisis porque se desprecian los efectos de reflexión y difracción que se producen en las lentes.

El modelo de dicha cámara describe la relación matemática entre las coordenadas de un punto 3D y su proyección sobre el plano de la imagen. Esto se basa en lo que se conoce como geometría proyectiva, cuyos principales axiomas son, según [4], los siguientes:

- Dos puntos definen una recta.
- Todo par de rectas se cortan en un punto. En caso de que estas dos rectas sean paralelas diremos que se cortan en el infinito.

Basándonos en estos dos axiomas podemos decir que la proyección de un punto de la escena se obtiene de la intersección que forma una línea que pasa por dicho punto y por el centro de proyección con el plano de la imagen.

Para describir y entender el proceso de captación de una escena por una cámara pin-hole es necesario introducir la geometría que compone dicho modelo, para ello nos ayudaremos de los textos anteriores ([1], [2], [3], [4]) y de la Figura 2.1. En ella podemos observar que el modelo tiene un centro óptico, C , contenido en el plano focal, por este punto pasarán todos los rayos de la proyección. Esos rayos tienen como finalidad encontrar el plano de la imagen para formar la llamada imagen proyectada. La distancia f que separa ambos planos se denomina distancia focal.

Como explica Faugeras en [2], para formar la proyección o imagen proyectada es necesario hacer una serie de transformaciones y referirse a los diferentes sistemas de coordenadas que pueden verse en la Figura 2.1:

- Sistema de coordenadas del mundo (X, Y, Z): la elección de estas coordenadas es arbitraria y con ellas se describe la posición del punto M en 3D respecto de la escena.

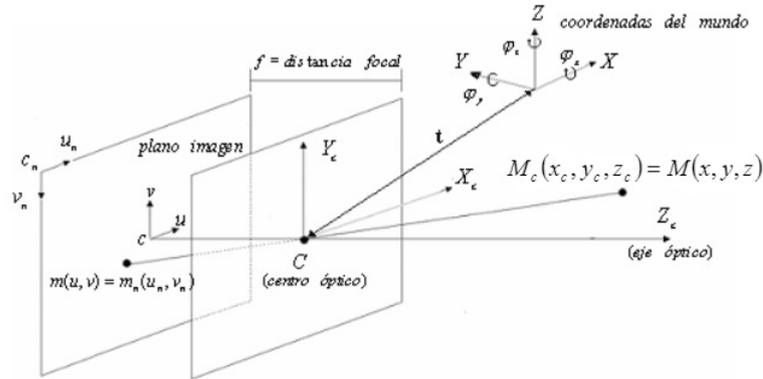


Figura 2.1: Modelo de cámara pin-hole [4].

- Sistema de coordenadas de la cámara (X_c, Y_c, Z_c) : son las coordenadas que describen la posición del punto M_c respecto a la cámara. El origen de este sistema de coordenadas es el centro óptico, C , de la cámara.
- Sistema de coordenadas de la imagen (u, v) : describen la posición del punto m en 2D con respecto al plano de la imagen. Este sistema de coordenadas tiene como origen el punto c , es decir donde el eje óptico corta al plano de la imagen.
- Sistema de coordenadas normalizadas de la imagen (u_n, v_n) : describen la posición del punto m_n en 2D con respecto al plano de la imagen. Lo único que hace a este sistema de coordenadas diferente al anterior es que tiene como origen de coordenadas la esquina superior izquierda del plano de la imagen (c_n) .

Tenemos que resaltar, que el punto M es idéntico al punto M_c , al igual que ocurre con el punto m y m_n , pero nos referimos a ellos con diferente notación para identificarlos con el sistema de coordenadas adecuado.

Además, todas las cámaras tienen una serie de parámetros que hacen referencia a este modelo de proyección, dichos parámetros podemos clasificarlos en dos grupos:

1. **Parámetros extrínsecos:** definen la orientación y la posición de la cámara con respecto a un sistema de coordenadas conocido (Sistema de coordenadas del mundo, Figura 2.1). Se compone de:
 - Traslación: es el vector que mide la distancia entre el centro óptico de la cámara (C) y el centro de coordenadas del origen del mundo. Según la Figura 2.1 este vector corresponde con $t = [t_x, t_y, t_z]^T$.

- Rotación: se trata de los ángulos rotados sobre cada uno de los ejes del sistema de coordenadas de la imagen con respecto al sistema de coordenadas del mundo. $[\varphi_x, \varphi_y, \varphi_z]$.

2. **Parámetros intrínsecos:** Describen la óptica y geometría interna de la cámara. Son los que afectan al proceso que sufre un rayo de luz desde que alcanza la lente de la cámara hasta que llega al plano de la imagen dónde debe ser proyectado. Estos parámetros son constantes para cada cámara pero diferentes para diferentes cámaras, y son:

- Centro del eje óptico, centro de la imagen o punto principal ($c = (u_0, v_0)$). Se trata del punto del plano de la imagen que corta con el eje óptico (Z_c).
- Factores de escala (k_u, k_v): relacionan el tamaño del objeto visto en la realidad con respecto a su tamaño una vez proyectado sobre plano de la imagen. Esta relación de proporcionalidad puede ser diferente en cada eje, por eso se compone de dos parámetros. A su vez, este factor de escala depende de:
 - ✧ Distancia focal (f): es la distancia existente entre el centro óptico y el centro del plano de la imagen.
 - ✧ Factores de conversión pixel-milímetros (d_u, d_v): se trata del número de píxeles por milímetro que usa la cámara. Se obtiene de la división del tamaño de la imagen en píxeles y el tamaño en milímetros del sensor CCD.
 - ✧ Factor de proporción (s): indica la relación de tamaño entre la dimensión horizontal y la vertical de un pixel.

La relación entre estos tres últimos parámetros para formar los factores de escala finales viene dada por:

$$k_u = s \cdot d_u \cdot f \qquad k_v = s \cdot d_v \cdot f \qquad (2.1)$$

Dicha relación únicamente es válida para este modelo ideal es decir, cuando no existe distorsión.

2.1.2. Matriz de Proyección

Toda cámara se modela mediante su matriz de proyección o matriz P , de tamaño 3×4 , dependiente de la matriz de parámetros intrínsecos K , que engloba a los

cinco componentes descritos en el apartado anterior, y las matrices de parámetros extrínsecos: rotación R y traslación \vec{t} .

Si suponemos que la cámara capta una escena y la proyecta sobre el plano de la imagen, es necesario encontrar la relación entre el punto M de la escena 3D y su proyección m sobre el plano de la imagen en 2D, dicha relación viene dada por:

$$m = P \cdot M \quad (2.2)$$

Esta igualdad es correcta excepto por un factor de escala:

$$\lambda m = P \cdot M \quad (2.3)$$

Para simplificar la definición de la transformación de proyección anterior de 3D a 2D se utilizan coordenadas homogéneas: si consideramos un punto 2D dado en coordenadas cartesianas $m = [x, y]$ y los expresamos en coordenadas homogéneas, dicho punto tendrá tres coordenadas: $m = [\alpha u, \alpha v, \alpha]$, donde las coordenadas cartesianas podrían recuperarse haciendo $\alpha = 1$. De manera similar, el punto 3D tendrá cuatro coordenadas en lugar de tres: $M = [\beta X, \beta Y, \beta Z, \beta]$.

Como hemos adelantado anteriormente, para determinar la forma de la matriz de proyección de una cámara es necesario realizar las transformaciones de los sistemas de coordenadas vistos en el apartado anterior. En el Apéndice A podemos encontrar explicadas todas estas transformaciones siguiendo a Faugeras en [2], aquí únicamente nos centraremos en aquella que tiene que ver con los parámetros extrínsecos que se quieren estimar, puesto que son éstos el objeto de análisis del trabajo de fin de grado que se describe. Terminaremos con la forma genérica de la matriz de proyección P .

Paso del sistema de coordenadas del mundo (X, Y, Z) al sistema de coordenadas de la cámara (X_c, Y_c, Z_c) :

En este caso, se trata de realizar una transformación Euclídea incluyendo los parámetros extrínsecos de rotación y traslación como la que nos muestra la Figura 2.2.

Este cambio de coordenadas que incluye tanto rotación como traslación, quedaría de la siguiente forma:

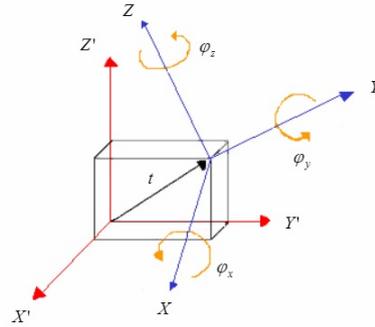


Figura 2.2: *Transformación Euclídea en 3D [4].*

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + t \quad (2.4)$$

En coordenadas homogéneas tendríamos (con $\beta = 1$):

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2.5)$$

Para llegar a la ecuación anterior vamos a centrarnos primero en la matriz de rotación R ; dicha matriz puede ser descompuesta en una multiplicación de las matrices de rotación sobre cada uno de los ejes, obteniendo: $R = R_x \cdot R_y \cdot R_z$. Ayudándonos de la Figura 2.3 podemos obtener dichas matrices de rotación individuales:

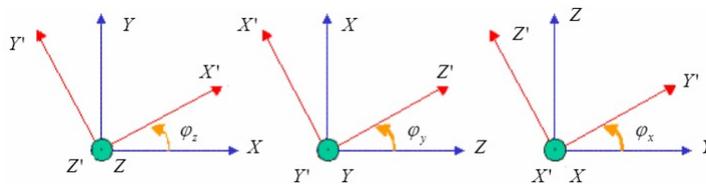


Figura 2.3: *Rotación de los ejes individualmente[4].*

$$R_z = \begin{bmatrix} \cos \varphi_z & \sin \varphi_z & 0 \\ -\sin \varphi_z & \cos \varphi_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.6)$$

$$R_y = \begin{bmatrix} \cos \varphi_y & 0 & -\sin \varphi_y \\ 0 & 1 & 0 \\ \sin \varphi_y & 0 & \cos \varphi_y \end{bmatrix} \quad (2.7)$$

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \varphi_x & \sin \varphi_x \\ 0 & -\sin \varphi_x & \cos \varphi_x \end{bmatrix} \quad (2.8)$$

Por lo tanto, la multiplicación de las matrices de rotación individuales dará la matriz R . Aplicamos esta matriz al sistema de coordenadas del mundo, para obtener un sistema intermedio (X_r, Y_r, Z_r) :

$$\begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad M_r = R \cdot M \quad (2.9)$$

Y, ahora, aplicamos la traslación sobre este sistema intermedio para llegar al sistema de coordenadas de la cámara:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad M_c = M_r + t \quad (2.10)$$

Si combinamos las Ecuaciones 2.9 y 2.10 conseguimos formular el paso directo de coordenadas del mundo a coordenadas de la cámara: $M_c = R \cdot M + t$, y pasando a coordenadas homogéneas se deduce la Ecuación 2.5.

Finalmente tenemos toda la información necesaria para saber que la matriz de proyección de una cámara se forma multiplicando la matriz de parámetros intrínsecos, K (Ecuación A.7 que podemos encontrar en el Apéndice A), y la matriz de parámetros extrínsecos de la Ecuación 2.5:

$$P = \begin{bmatrix} f \cdot d_u & 0 & u_0 & 0 \\ 0 & f \cdot d_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad P = K \cdot [R | t] \quad (2.11)$$

2.1.3. Geometría Epipolar y Matriz Fundamental

Según Hartley en [5], la geometría epipolar es la geometría proyectiva intrínseca entre dos vistas, es completamente independiente de la estructura de la escena y sólo depende de los parámetros intrínsecos de ambas cámaras así como de la posición relativa en la que se encuentran. En otras palabras, determina una serie de restricciones en la proyección de un punto del espacio sobre los planos de imagen de un par de cámaras. Estas restricciones las engloba lo que se conoce como *matriz fundamental*.

Para entender la geometría epipolar vamos a seguir las explicaciones de varios textos, entre ellos [5] y [6]. Considerando las dos cámaras que se muestran en la Figura 2.4a, observamos que el punto M , o punto 3D, está contenido en el mismo plano que los centros de proyección, C y C' , de las cámaras A y B respectivamente. A dicho plano se le denomina plano epipolar. Por otro lado, la línea que une los centros de las cámaras se denomina línea de base y corta a las dos imágenes por lo que se conoce como epipolos (e y e' en la Figura 2.4b). Todos los planos que pasan por los centros ópticos de las cámaras y cualquier punto M cortan a los planos de imagen en los epipolos, esta condición es conocida como condición de coplanaridad y es la base de la que deriva la geometría epipolar.

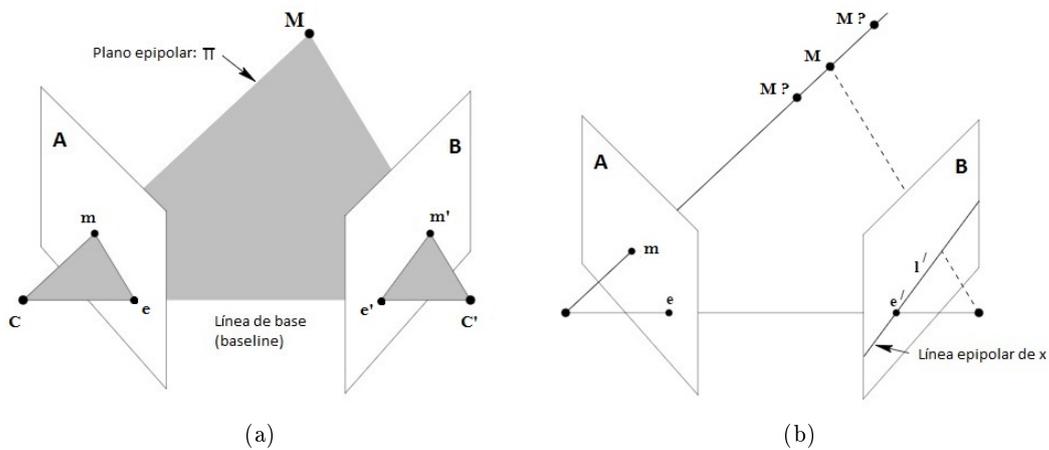


Figura 2.4: Geometría epipolar [5].

Vemos que el punto M se proyecta en el plano de la imagen A como m , de la misma forma, M se proyecta sobre la imagen B como m' y, dada esta geometría, observamos en la Figura 2.4b que el punto m' forma parte de la línea l' (denominada línea epipolar). La restricción epipolar derivada de esto establece lo siguiente: dado

un punto M , cuya proyección sobre el plano de la imagen A es m , tenemos que, el punto m' correspondiente a la proyección de M sobre el plano de la imagen B , deberá estar sobre la línea epipolar l' (ver Figura 2.4b), y viceversa, pudiendo así reducir considerablemente el espacio de búsqueda de correspondencias entre ambas imágenes. Es decir, la línea epipolar es la proyección en la segunda imagen del rayo desde el punto m a través del centro C de la primera cámara.

Como hemos adelantado al inicio de esta sección, y se explica en [5], la matriz fundamental (F) es la representación algebraica de la geometría epipolar siendo la encargada de relacionar los puntos de una imagen con las correspondientes líneas epipolares en la otra imagen, y lo hace de la siguiente forma:

$$l' = F \cdot m \quad (2.12)$$

Hay que destacar, que para cualquier par de puntos correspondientes entre las dos imágenes, la matriz fundamental cumple: $m' \cdot F \cdot m = 0$, por lo tanto es posible calcular F a partir de correspondencias entre ambas imágenes.

Propiedades de la matriz fundamental (según Hartley en [5])

- Traspuesta: Si F es la matriz fundamental asociada a las cámaras con matrices de proyección P y P' , entonces F^T es la asociada a P' y P .
- Líneas epipolares: Para el punto m de la Figura 2.4b, la línea epipolar asociada viene dada por la Ecuación 2.12. De forma inversa, para el punto m' , la línea epipolar asociada es $l = F^T \cdot m'$.
- Grados de libertad: F es una matriz de tamaño 3×3 y tiene sólo 7 grados de libertad ya que $\det(F) = 0$.

Estimación de la matriz fundamental a partir de correspondencias

Si conocemos los parámetros de ambas cámaras, es decir, sus matrices de proyección, entonces tenemos información suficiente para obtener la geometría epipolar. Generalmente dichos parámetros no se conocen, por lo que existen numerosos métodos para la estimación de la matriz fundamental a partir de correspondencias entre las imágenes capturadas por las dos cámaras. En general, dado un conjunto de correspondencias tentativas entre dos imágenes, debido a la existencia de correspondencias ruidosas e incorrectas, no es posible estimar directamente la geometría epipolar de la escena. Por ello, los métodos existentes proponen estrategias para reducir la influencia

de las correspondencias incorrectas (conocidas como *outliers*), dichos métodos pueden clasificarse en: métodos lineales, iterativos o robustos.

Siguiendo [7], todos estos métodos están basados en la obtención de una estimación de la matriz fundamental mediante la resolución de un sistema de ecuaciones que se deduce de $m' \cdot F \cdot m = 0$. Este sistema de ecuaciones puede escribirse como:

$$U \cdot f = 0 \quad (2.13)$$

donde

$$f = [F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33}]^T \quad (2.14)$$

$$U = \begin{pmatrix} u_1 u'_1 & u_1 v'_1 & u_1 & v_1 u'_1 & v_1 v'_1 & v_1 & u'_1 & v'_1 & 1 \\ \vdots & \vdots \\ u_n u'_n & u_n v'_n & u_n & v_n u'_n & v_n v'_n & v_n & u'_n & v'_n & 1 \end{pmatrix} \quad (2.15)$$

F_{ij} corresponde al elemento de la fila i y de la columna j de la matriz fundamental.

Es importante notar que hay sólo siete parámetros independiente y nueve incógnitas. Los siete parámetros independientes vienen dados por las dos columnas independientes de la matriz fundamental y el factor de escala. Esta configuración fuerza a la matriz fundamental a ser de rango 2.

- **Métodos lineales.** El *método lineal de siete puntos* esta basado en el cálculo de la matriz fundamental usando únicamente siete correspondencias. Debido a la homogeneidad de las ecuaciones anteriores, la solución viene dada por un conjunto de matrices de la forma: $F = \alpha \cdot F_1 + (1 - \alpha) \cdot F_2$. Forzando al rango de la matriz a ser igual a 2 y usando la expresión $\det(\alpha \cdot F_1 + (1 - \alpha) \cdot F_2)$, se obtendría un polinomio que tiene que ser resuelto para la obtención de α y F . La gran ventaja de este método es que la matriz fundamental puede ser estimada sólo con siete correspondencias. Esto se convierte en un inconveniente cuando dichos puntos están mal localizados, además, este método no puede aplicarse en presencia de redundancia, es decir, con $n > 7$. Otro método interesante es el *método de ocho puntos*, donde la redundancia de puntos permite la minimización del error en la estimación de F mediante $\min_F \sum (m'_i \cdot F \cdot m_i)^2$. El método clásico para resolver una ecuación de este tipo es la técnica de *mínimos cuadrados* obligando a uno de los componentes de F a ser 1.
- **Métodos iterativos.** Los métodos iterativos pueden clasificarse en dos grupos: aquellos que minimizan las distancias (d) entre puntos y líneas epipolares y los

basados en gradiente. Los primeros, tratan de minimizar la siguiente ecuación:

$$\min_F \sum_i (d^2(m_i, F^T m'_i) + d^2(m'_i, F m_i)) \quad (2.16)$$

Una primera aproximación consiste en aplicar métodos iterativos como Newton-Raphson o Levenberg-Marquardt en la ecuación anterior. Otra aproximación propone reescribir la ecuación 2.16 como $\min_F \sum w_i^2 (m_i^T \cdot F \cdot m'_i)^2$ donde $F^T m'_i = (l_1, l_2, l_3)^T$, $F m_i = (l'_1, l'_2, l'_3)^T$ y $w_i = (\frac{1}{l_1^2 + l_2^2} + \frac{1}{l'_1{}^2 + l'_2{}^2})^{1/2}$. Este método iterativo consiste en calcular el valor del peso w_i equivalente a las distancias epipolares mediante el uso de la F de la iteración anterior (en la primera iteración $w_i = 1$) y luego minimizar mediante el uso de mínimos cuadrados en cada iteración.

El segundo grupo basado en gradiente, trata de resolver la siguiente ecuación:

$$\min_F \sum_i (m_i^T F m'_i)^2 / g_i^2 \quad (2.17)$$

donde $g_i = \sqrt{l_1^2 + l_2^2 + l'_1{}^2 + l'_2{}^2}$.

- **Métodos robustos.** Entre los que se encuentra RANSAC. Se trata de un algoritmo aleatorio e iterativo que sirve para calcular un modelo matemático sobre un conjunto de datos (en este caso correspondencias) que contienen valores atípicos. Consiste en la selección de un subconjunto aleatorio de las correspondencias encontradas y, a partir de ellas, generar un modelo. Una vez hecho esto, se prueban el resto de datos sobre el modelo generado. Dicho modelo se considerará bueno si ha clasificado suficientes puntos como parte de lo que se conoce como conjunto de consenso, de ser así se pasaría a la fase de mejora en la cual se utiliza dicho conjunto para reestimar el modelo matemático. Si el modelo matemático no es bueno se vuelve a elegir un conjunto aleatorio de correspondencias iniciales y se repiten todos los pasos. Por último, además de una estimación de la matriz fundamental, RANSAC también devolverá el conjunto de *inliers* con los que ha realizado dicha estimación.

Este tipo de métodos englobaría también algoritmos como M-Estimators, LMedS (Least-Median-Squares), MLESAC y MAPSAC.

Medidas de error

Para saber cómo de válida es nuestra estimación inicial y cuál es la diferencia entre las diferentes aproximaciones existentes para su obtención, es necesario establecer

una medida de error de la matriz fundamental. En [8] se exponen principalmente dos formas de medida de error: siendo F la estimación de la matriz fundamental tenemos, por un lado, el error geométrico, calculado como el error cuadrático medio de la distancia entre las correspondencias que estiman dicha matriz ($m \longleftrightarrow m'$) y la línea epipolar que generan:

$$+ \quad \text{ErrorGeom}(F) = \sqrt{\frac{1}{n} \cdot \sum d(m, Fm')^2} \quad (2.18)$$

Y, por otro lado, lo que los autores denominan error real. Para medirlo, se genera una nube de 1000 correspondencias correctas ($m^* \longleftrightarrow m'^*$) y se calcula el error obtenido para dichas correspondencias mediante la Ecuación 2.18.

Posición relativa de una cámara respecto de la otra

La idea en la que se basa nuestro trabajo es conseguir una estimación inicial de F y refinarla mediante diferentes técnicas relacionadas con la geometría epipolar. Una vez conseguida dicha estimación se extraerían las matrices de proyección, P y P' , asociadas a ambas imágenes y sería posible la obtención de los parámetros extrínsecos de rotación y traslación de una de las cámaras estableciendo la posición de la otra como referencia. Este proceso lleva implícito la ambigüedad introducida por el uso de la matriz fundamental. La matriz fundamental es independiente de la elección del origen del sistema de coordenadas del mundo y de la calibración interna de cada cámara. Por lo tanto, permanece inalterada ante una transformación proyectiva del espacio 3D. En otras palabras, sea H una transformación proyectiva del espacio 3D y P y P' las matrices de proyección asociadas a dos cámaras, la matriz fundamental para el par P y P' y el par PH y $P'H$ será la misma [5].

Para realizar este proceso seguiremos las indicaciones de [5]:

- Se define una de las matrices de proyección en su forma canónica:

$$P = \left[I | \vec{0} \right] \quad (2.19)$$

, donde I es la matriz identidad 3×3 y $\vec{0}$ es el vector de ceros 3×1 .

- La matriz de proyección de la otra cámara vendrá determinada por (salvo por una homografía de escena):

$$P' = \left[[e']_x F_f | e' \right] \quad (2.20)$$

, donde $[e']_x$ es la expansión en matriz antisimétrica de su epipolo.

- El vector de traslación \vec{t} y la matriz de rotación R' de C' respecto de C puede obtenerse a partir de P' :

$$R' = P'(:, 4) * (K')^{-1} / \|P'(:, 4) * (K')^{-1}\|_{fro} \quad (2.21)$$

$$\vec{t} = -P'(:, 1, 2, 3) \quad (2.22)$$

, donde $P'(:, j)$ es la submatriz resultante de eliminar la columna j , $\|, \|_{fro}$ es el operador que realiza la norma de Frobenius para matrices cuadradas y $(K')^{-1}$ es la inversa de la matriz de parámetros intrínsecos de la cámara analizada. El procedimiento para la obtención de la matriz de rotación R' puede verse como una descomposición QR en una matriz ortogonal R y otra triangular superior conocida K' .

2.2. Detección de puntos de interés

Como hemos visto, la correcta estimación de la matriz fundamental vendrá determinada por la calidad de las correspondencias encontradas. Normalmente, la búsqueda de correspondencias entre dos imágenes está compuesta de tres pasos: la detección de puntos de interés o puntos característicos, su descripción y la obtención de las correspondencias. En este apartado nos centraremos en la detección. Consiste en la extracción de puntos que, por sus características, son susceptibles a mantenerse iguales a pesar de los cambios en la imagen. Es decir, se trata de encontrar aquellos puntos que sean invariantes a transformaciones típicas como traslaciones, rotaciones, cambios de escala, de perspectiva o de iluminación.

2.2.1. Scale-Space

Consiste en la representación de una imagen como una familia de imágenes suavizadas mediante el uso de un kernel, normalmente Gaussiano porque no introduce nuevas estructuras a la imagen. Conforme aumenta el suavizado, aumenta lo que se conoce como *escala* (ψ). En el ejemplo de la Figura 2.5 observamos que, generando este scale-space, ciertas estructuras prevalecen o incluso aumentan su significancia con el aumento de escala. Dichas estructuras serán las que nos pueden ayudar a extraer los puntos característicos de la imagen, es decir, serán las que habrá que tener en cuenta a la hora de buscar correspondencias entre un par de imágenes.



Figura 2.5: *Ejemplo de Scale-Space.*

Basándonos en la explicación de Lindeberg en [9], si partimos de una imagen $I(u, v)$, su scale-space ($L(u, v; \psi)$) estará formado por la convolución de dicha imagen con un filtro Gaussiano definido como:

$$g(u, v; \psi) = \frac{1}{2 \cdot \pi \cdot \psi} e^{-(u^2+v^2)/2\psi} \quad (2.23)$$

Se utiliza este tipo de filtro porque no introduce nuevas estructuras a la imagen. De forma que, obtendríamos: $L(u, v; \psi) = g(u, v; \psi) * I(u, v)$

La escala ψ suele ser establecida como proporcional a la varianza del filtro Gaussiano: $\psi_i = i\sigma^2$. Para $i = 0$, no existe suavizado, siendo la primera imagen del scale-space la imagen original como veíamos en la Figura 2.5.

Utilizar un detector sobre este espacio hará que los puntos extraídos sean invariantes a escala, esto es necesario para hacer frente a las posibles variaciones de tamaño sobre las imágenes.

2.2.2. Detectores sobre el Scale-Space

Existen numerosos detectores sobre el Scale-Space basados principalmente en las derivadas de la señal. Además si dichas derivadas se normalizan es posible obtener la escala a la que se producen las características. Algunos de estos detectores utilizan la Laplaciana de la Gaussiana (LoG), la diferencia de Gaussianas (DoG) o el Determinante del Hessiano (DoH).

- **LoG.** Siendo $L(u, v; \psi)$ la representación del Scale-Space Gaussiano, la Laplaciana normalizada respecto de la escala viene dada por la expresión $\nabla^2 L = L_{uu} + L_{vv}$, donde L_{uu} y L_{vv} son las segundas derivadas. Con esto obtendremos máximos/mínimos relativos en el volumen $(u, v; \psi)$ cuando existan blobs oscuros/claros permitiendo localizar tanto el blob como la escala a la que se detecta.
- **DoG.** En este caso la detección se realizará sobre la resta de dos Gaussianas consecutivas $D(u, v; \psi) = L(u, v; \psi_{i+1}) - L(u, v; \psi_i)$. De nuevo se obtendrán los máximos y mínimos relativos en el volumen generado para obtener los puntos de interés.
- **DoH.** La matriz Hessiana (HL) viene dada por las segundas derivadas de la función, es decir, describe la curvatura. Se denomina Hessiano al determinante de dicha matriz. La detección de puntos sobre el espacio generado por este determinante también es muy característica y la utilizan detectores como D_{1norm} que se ha desarrollado en este trabajo y que explicaremos en la Sección 3.2.1.

$$HL(u, v; \psi) = \begin{bmatrix} L_{uu} & L_{uv} \\ L_{vu} & L_{vv} \end{bmatrix} \quad \det HL(u, v; \psi) = L_{uu} \cdot L_{vv} - L_{uv}^2 \quad (2.24)$$

- **SIFT.** Es una versión optimizada computacionalmente del detector LoG, utiliza pirámides (reducción progresiva del tamaño de la imagen) para reducir el número de suavizados necesarios y así abaratar el coste computacional.
- **SURF.** Se trata de una versión optimizada computacionalmente del detector DoG, utiliza filtros de caja de tamaño creciente para simular el incremento de escala y abaratar así el coste computacional.

2.3. Descripción de puntos de interés

Una vez localizados los puntos de interés que servirán para establecer las correspondencias, pasaríamos a la fase de descripción. El objetivo de esta fase es conseguir, para cada punto, una información distintiva y relevante de la región que lo rodea, de esta forma la misma región podrá ser reconocida en otra imagen. De nuevo, se trata de conseguir una descripción invariante a las transformaciones típicas enumeradas anteriormente.

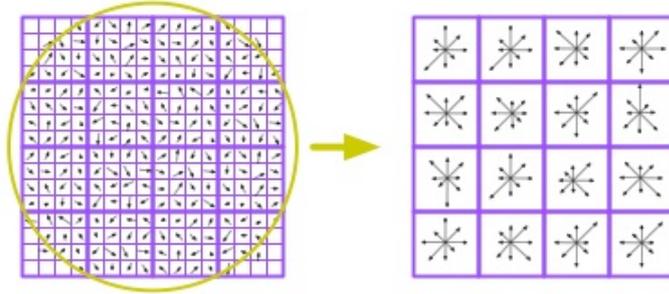


Figura 2.6: Orientaciones del descriptor SIFT.

2.3.1. SIFT (Scale-Invariant Feature Transform) y DSIFT (Dense Scale-Invariant Feature Transform)

El descriptor SIFT fue presentado por Lowe [10] a la vez que el detector. Tanto SIFT como su versión para descripción densa (Dense-SIFT o DSIFT) se calculan sobre regiones locales, generalmente, centradas en puntos de interés. Cada uno de los píxeles pertenecientes a esta región local contribuye a un histograma, dependiendo de su ubicación en la región, y de la orientación y la norma del gradiente de la imagen en dicha localización. Para calcular el descriptor debemos conocer la escala a la que ha sido detectado el punto de interés y trabajar sobre la imagen suavizada, de esta forma el descriptor será invariante a escala. El paso principal es la obtención del gradiente de la imagen sobre la que trabajamos así como su orientación. Una vez hecho esto, se genera un histograma de orientaciones de 36 bins (1 orientación cada $360^\circ/36 \text{ bins} = 10^\circ$) sobre el cual, cada pixel de la región local vota en función de la magnitud (ponderada con un filtro Gaussiano centrado en el punto a describir) y la orientación de su gradiente (que se rota respecto a la orientación dominante del punto de interés). De este histograma se extraen las orientaciones principales haciendo al descriptor invariante a rotaciones.

En la Figura 2.6 vemos que se describe una región de 16×16 píxeles entorno al punto de interés. A su vez, esta región se divide en subregiones de 4×4 píxeles obteniendo un histograma de orientaciones de 8 bins (1 orientación cada $360^\circ/8 = 45^\circ$) para cada subregión, de esta forma, se consigue que el descriptor sea robusto a desplazamientos locales. Por último, el descriptor se extrae de la concatenación de los 16 histogramas generados por las subregiones anteriores.

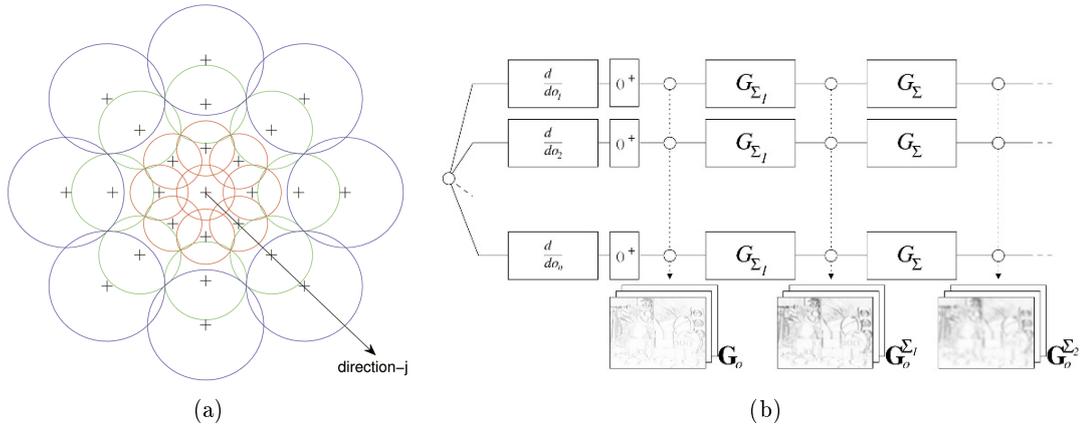


Figura 2.7: DAISY. En la (a) podemos ver la forma del descriptor DAISY. Y en (b) observamos cómo se calcula dicho descriptor. Ambas Figuras provienen de [11].

2.3.2. DAISY

Según los autores de [11], se trata de un descriptor local de imagen inspirado en SIFT y GLOH. El objetivo de este descriptor es reformular los dos anteriores para que puedan ser calculados de forma eficiente cuando se quiere describir un gran número de puntos o, incluso, imágenes completas. Intuitivamente, dicho objetivo se conseguirá calculando los histogramas de orientaciones una única vez y volviéndolos a utilizar para todos los píxeles vecinos. Con este fin y como podemos ver en la Figura 2.7a, DAISY reemplaza las sumas ponderadas de las normas de los gradientes por circunvoluciones de los gradientes en direcciones específicas mediante el uso de varios filtros Gaussianos.

Analicemos el parecido entre SIFT y DAISY para entender éste último. Cada vector de gradiente que obteníamos en SIFT para el cálculo de un solo descriptor, ahora con DAISY contribuirá a varios de los elementos del vector de descripción, pero en este caso, la suma de las contribuciones no se realiza mediante ponderación como ocurría con SIFT sino que es calculada a través de convoluciones para mayor eficiencia computacional. En primer lugar, se calculan los mapas de orientaciones de las imágenes originales como podemos ver en la Figura 2.7b, estos mapas luego serán convolucionados para obtener los mapas de orientación convolucionada, $G_o^{\Sigma_i}$. El valor de $G_o^{\Sigma_i}$ se corresponde conceptualmente con los valores de los bins de SIFT, y serán usados para la construcción del descriptor DAISY.

Para hacer que DAISY sea invariante a rotaciones estableceremos las direcciones de los gradientes en cada uno de los puntos tomando como referencia la dirección

dada por la línea epipolar sobre el punto bajo análisis.

2.4. Correspondencias entre imágenes

En esta Sección se explicarán las posibles formas de extracción de correspondencias. En este caso evaluaremos dos métodos diferentes, mediante el uso de descriptores (2.4.1) o mediante mapas de confianza de correspondencias densas (2.4.2).

2.4.1. Método de obtención de correspondencias

Según Lowe [10], para establecer correspondencias entre puntos de interés descritos mediante un vector de características, únicamente hay que calcular la distancia Euclídea entre descriptores de puntos en la imagen A y puntos en la imagen B mediante la Ecuación 2.25. Es decir, se define como correspondencia para un punto m en la imagen A aquel punto m' en la imagen B cuyo descriptor tengan menor distancia, es decir, sea más parecidos al descriptor de m . Sea $desc()$ el operador que extrae la descripción (e.g. DAISY o SIFT) para un punto en la imagen. La distancia entre dos puntos vendrá determinada por:

$$d(m, m') = \sqrt{\sum (desc(m) - desc(m'))^2} \quad (2.25)$$

Por último, de entre estas correspondencias, para eliminar correspondencias ruidosas, se seleccionan solo aquellas que generen una distancia de correspondencia que sea, el menos, η veces menor que la obtenida para cualquier otro punto de la imagen B . Donde η es establecido por el usuario: a mayor η mayor restrictividad.

2.4.2. NRDC (Non-Rigid Dense Correspondence)

Este algoritmo está diseñado para el cálculo de un conjunto denso de correspondencias fiables entre dos imágenes que comparten algunos contenidos pero pueden diferir en varias condiciones de la escena o de la cámara. La región de correspondencias densas se consigue alineando pequeños parches transformados por geometría simple y transformaciones fotométricas para lograr la robustez a dichos cambios en las condiciones de la escena o las cámaras. Se encuentran coincidencias a través de cambios significativos de contenido o de pose, a la vez que se establece esta región de correspondencias, se diseña un modelo de transformación de color.

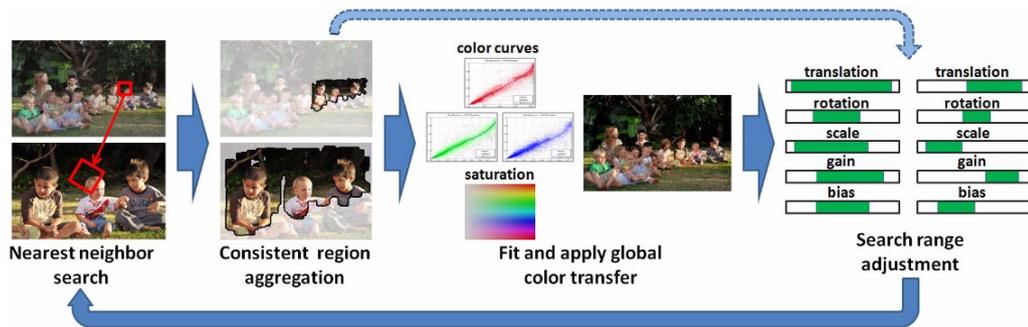


Figura 2.8: Descripción del algoritmo NRDC según [12].

En resumen, como se explica en [12] y podemos ver en la Figura 2.8, NRDC sigue los siguientes pasos para cada una de las escalas:

1. Búsqueda del vecino más próximo: para cada parche en la imagen origen encontrar su vecino más cercano en la imagen de referencia, esto se hace buscando en un rango limitado de traslaciones, escalas, rotaciones, etc.
2. Región de agregación: someter a las regiones de correspondencias encontradas a unas pruebas de consistencia, si dichas pruebas son superadas se considerarán correspondencias fiables.
3. Transformación de color apropiada: generar el modelo de transformación de color basado en las regiones de correspondencias fiables.
4. Ajustar el rango de búsqueda mediante las regiones de correspondencias y el modelo de transformación de color.

Este proceso se repetirá, iterativamente y hasta su convergencia, con el fin de perfeccionar el modelo.

Capítulo 3

Diseño y desarrollo

3.1. Introducción

En este capítulo describiremos los algoritmos desarrollados. Todos comparten un objetivo final: la obtención de una estimación de la matriz fundamental entre dos vistas a partir de la cual se puedan extraer las matrices de proyección de ambas cámaras y, por consiguiente, la posición de una con respecto a la otra. Para ello seguiremos los pasos de la Figura 3.1 a partir de los cuales llegaremos a una primera estimación de dicha matriz fundamental y posteriormente procederemos a una fase de refinamiento con el fin de mejorarla para terminar con el cálculo de los parámetros extrínsecos, R y \vec{t} .

3.2. Búsqueda de correspondencias iniciales

Partiendo de dos puntos de vista diferentes de una misma escena, en adelante, y respetando la nomenclatura introducida en el estado del arte, imagen A e imagen B ,



Figura 3.1: *Diagrama de bloques.*

el objetivo de esta sección es encontrar las correspondencias existentes entre dichas vistas para la posterior estimación de la matriz fundamental. Para conseguir esto se han utilizado dos métodos diferentes que explicaremos a continuación. Hay que destacar que esta búsqueda de correspondencias inicial únicamente se calcula una vez para ahorrar coste computacional al algoritmo.

3.2.1. Correspondencias locales

Este método engloba los pasos de detección y descripción de puntos así como la posterior búsqueda de correspondencias.

Basándonos en el texto [13] de Lindeberg, hemos implementado un detector de puntos de interés sobre el Scale-Space que responde, en cada pixel de la imagen, a la fuerza de la característica Hessiana. Para ello, es necesario el cálculo del determinante normalizado de la matriz Hessiana (Hessiano) y de su traza (Ecuación 2.24) para todos los píxeles de la imagen, a partir de sus segundas derivadas:

$$\det H_{norm}L = \psi^2 (L_{uu} \cdot L_{vv} - L_{uv}^2) \quad traceHL = L_{uu} + L_{vv} \quad (3.1)$$

Posteriormente aplicamos la fórmula del detector descrito por Lindeberg ([13]) que incluye dos variantes. Detector sin signo y detector con signo:

$$D_{1,norm}L = \begin{cases} \psi^2(\det HL - k \cdot trace^2 HL) & si \det HL - k \cdot trace^2 HL > 0 \\ 0 & \end{cases} \quad (3.2)$$

$$D_{1,norm}L = \begin{cases} \psi^2(\det HL - k \cdot trace^2 HL) & si \det HL - k \cdot trace^2 HL > 0 \\ \psi^2(\det HL + k \cdot trace^2 HL) & si \det HL + k \cdot trace^2 HL < 0 \\ 0 & \end{cases} \quad (3.3)$$

Donde HL es la matriz Hessiana calculada en cada punto de la imagen, ψ es la escala actual sobre la que se extraen los puntos de interés y k es una constante de valor 0.06 establecido por [13].

De los resultados derivados de este detector nos quedamos únicamente con cierto número de puntos. En el caso del detector con signo dichos puntos serán elegidos como los mínimos y máximos, locales y entre escalas, más fuerte de entre los valores derivados del detector. Si el detector es sin signo, nos quedaremos con los máximos más fuertes. Además, para evitar que las correspondencias se agrupen únicamente en

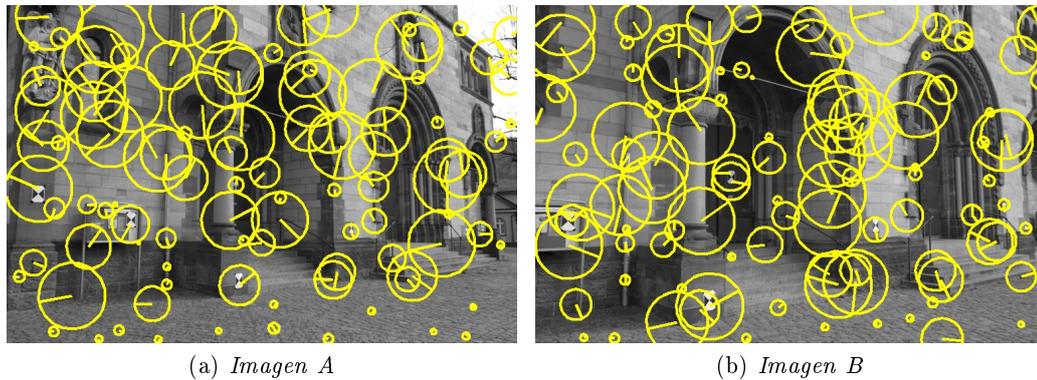


Figura 3.2: *Ejemplo de detección de puntos sobre el Scale-Space con descripción SIFT. Únicamente se muestran 100 puntos de interés*

una parte del plano de la imagen se ha obligado, mediante el uso de una rejilla, a que exista mínimo un punto de interés en cada una de las divisiones de la imagen. Esto se hace con el fin de que la estimación de la matriz fundamental tenga en cuenta toda la imagen y no solo parte de ella.

Una vez seleccionados los puntos característicos se procedería a su descripción mediante el descriptor SIFT (Sección 2.3.1). En la Figura 3.2 podemos ver un ejemplo de los puntos detectados sobre el Scale-Space así como de la orientación dominante del gradiente en la descripción SIFT. El tamaño del radio de las circunferencias es equivalente a la escala a la que ha sido detectado el punto de interés.

Por último y como ya se explicó en la Sección 2.4, una vez obtenidos los puntos de interés y sus descripciones buscamos las correspondencias midiendo el parecido entre los descriptores de ambas imágenes a partir de la distancia Euclídea existente entre ellos y seleccionando aquellas correspondencias que resulten en distancias mínimas con puntos de la otra imagen η veces menores que la segunda menor distancia obtenida.

Como hemos adelantado, y podemos intuir con la Figura 3.2, existen varios parámetros de configuración para la búsqueda de correspondencias locales mediante este método como, por ejemplo, el número de puntos de interés, el número de escalas del Scale-Space o, incluso, la forma del detector (con signo o sin signo). Por lo tanto, para generar una estimación inicial adecuada hemos realizado un barrido de los parámetros de configuración, es decir, del número de puntos y del número de escalas, así nos quedaremos con la configuración que produzca un menor error de F .

3.2.2. Correspondencias densas

El otro método utilizado para la búsqueda de correspondencias es NRDC debido a su gran capacidad para encontrar correspondencias densas entre dos imágenes (como ya se describió en la Sección 2.4.2). En este caso, la búsqueda de correspondencias no necesita de una detección previa de puntos de interés, sino que NRDC genera unos mapas de confianza densos mediante los cuales relaciona las imágenes A y B .

En la aproximación diseñada se propone extraer todas las correspondencias cuya confianza fuese igual a 1 obteniendo así un conjunto elevado de correspondencias. Debido al gran tamaño de este conjunto, y con el objetivo de reducir coste computacional del algoritmo, nos hemos visto obligados a hacer una preselección de correspondencias antes de estimar la matriz fundamental. Esta preselección se ha realizado calculando los descriptores DAISY de los puntos correspondientes en ambas imágenes, midiendo la distancia Euclídea entre ellos y seleccionando las N correspondencias que resultan en menores distancias.

3.3. Estimación de la matriz fundamental inicial

Una vez obtenidas las correspondencias entre ambas imágenes se procede al cálculo de una primera estimación de la matriz fundamental. Esta estimación es obtenida introduciendo dichas correspondencias al algoritmo RANSAC (explicado en la Sección 2.1.3). El resultado de este proceso deriva en la obtención de una matriz fundamental inicial, que denominaremos F_0 .

3.4. Refinamiento

Una vez conseguida la matriz fundamental inicial (F_0), trataremos de mejorar dicha estimación por medio de un conjunto de estrategias diseñadas en el transcurso de este trabajo de fin de grado. En esta sección explicaremos estas técnicas basadas, principalmente, en el cambio en la descripción de los puntos y en la mejora de las correspondencias que estimaban inicialmente la matriz fundamental.

3.4.1. Mejora de la descripción

En primer lugar, para este proceso de refinamiento, es necesario elegir un método de descripción densa de puntos. Hablamos de descripción densa porque las técnicas de mejora de correspondencias que se explicarán a continuación están basadas en un análisis exhaustivo de todos los puntos existentes sobre las rectas epipolares o sobre

las máscaras que rodean dichas rectas, por lo que calcular todos los descriptores de las imágenes en un solo paso evitará la necesidad de repetir su cálculo. Además, de esta forma evitamos la influencia de un detector de singularidades preliminar, que podría limitar el número de puntos a analizar, examinando así todos los posibles candidatos a ser coincidentes con el punto inicial.

Los descriptores elegidos son DAISY (explicado en la Sección 2.3.2) y DSIFT que, como ya veíamos en la Sección 2.3.1, es exactamente igual a SIFT pero desarrollado para describir una gran cantidad de puntos.

3.4.2. Mejora de las correspondencias

Una vez se disponga de los descriptores densos de las imágenes A y B se pasa a la mejora de las correspondencias que inicialmente estimaban F_0 , es decir, las devueltas por el algoritmo RANSAC tras la fase de búsqueda de correspondencias locales o densas. Para ello hemos implementado tres métodos diferentes basados en la restricción epipolar que mencionábamos en la Sección 2.1.3 del Estado del Arte:

“Dado un punto M , cuya proyección sobre el plano de la imagen A es m , tenemos que, el punto m' correspondiente a la proyección de M sobre el plano de la imagen B , deberá estar sobre la línea epipolar l' , y viceversa”

3.4.2.1. Proyección y Retroproyección

El primero de los métodos implementados, denominado *Proyección y retroproyección*, parte del supuesto de que la primera estimación de la matriz fundamental tiene un error bajo, es decir, es una estimación aceptable pero mejorable. Como vemos en la Figura 3.3, este método parte de un punto en la imagen A , m , con el cual se genera la línea epipolar que constriñe sus correspondencias en la imagen B mediante $l' = F_0 \cdot m$. Una vez hecho esto se analizan las distancias entre el descriptor del punto en A y los descriptores de los puntos que están sobre la línea epipolar en B ; de este paso se extrae el punto en B (m') cuyo descriptor es más parecido (en términos de distancia Euclídea) al descriptor del punto m . Denominamos a esta etapa: proceso de proyección.

Por otro lado, el proceso de retroproyección que se muestra en la Figura 3.4, parte del punto m' que se extrajo en el proceso de proyección. El número de potenciales correspondencias a m' en A se constriñe a puntos en la línea epipolar de m' : $l = F_0^T \cdot m'$. De entre estas potenciales correspondencias el punto de retroproyección m será aquel

Figura 3.3: *Proyección.*Figura 3.4: *Retroproyección.*

cuya descripción sea más parecida a la del punto m' , siempre en términos de distancia Euclídea.

Por último, si el punto de retroproyección, m^* , es igual al punto de partida en A , m , o bien se encuentra dentro del área de vecindad (definida mediante la Ecuación 3.4) que establece el radio dado por la escala a la que se detecta dicho punto de partida, se establecerá que la relación entre el punto inicial y el punto en B es correcta, es decir, son correspondientes, y por lo tanto m y m' también lo son.

$$(u_{m^*} - u_m)^2 + (v_{m^*} - v_m)^2 \leq (\psi_m)^2 \quad (3.4)$$

Mediante este proceso eliminamos correspondencias $m \leftrightarrow m'$ incorrectas e introducimos nuevas correspondencias $m^* \leftrightarrow m'$. Nótese, que si realizamos este procedimiento de manera iterativa, podríamos generar correspondencias donde ninguno de los puntos que las compusiesen fuesen parte de la correspondencia inicial: $m \leftrightarrow m'$. En este trabajo, se propone realizar un máximo de dos procesos de proyección con



Figura 3.5: *Búsqueda exhaustiva: proyección con máscara*

uno de retroproyección intercalado, es decir, pueden alcanzarse correspondencias de naturaleza: $m^* \longleftrightarrow m^{*'}$ siguiendo el procedimiento: proyección \rightarrow retroproyección \rightarrow proyección.

3.4.2.2. Búsqueda exhaustiva mediante máscara

En este segundo método no se tiene en cuenta el paso de retroproyección, es decir, el punto sobre la imagen B que se encuentre durante la fase de proyección será el que se asocie como correspondencia al punto de partida.

Para generar una técnica que realmente intente corregir el error de la estimación inicial de la matriz fundamental, F_0 , hemos ampliado el área de búsqueda alrededor de la línea epipolar generando una máscara a partir de los radios dados por las escalas de cada uno de los puntos que caen sobre dicha línea epipolar. Es decir, se permiten correspondencias fuera de la línea epipolar. Un ejemplo de esta situación puede observarse en la Figura 3.5. Dado un punto m en la imagen A se construye una máscara de candidatos potenciales alrededor de su línea epipolar en la imagen B . El proceso consiste en calcular la distancia entre el descriptor del punto en A y los descriptores de todos los puntos activados en la máscara sobre B (en la Figura 3.5 el área de aplicación de la máscara se ilustra por la zona que contiene información de la imagen) y quedarnos con aquel punto m' en B cuyo descriptor sea más parecido a m .

3.4.2.3. Combinación

Esta última técnica desarrollada está formada por la combinación de las dos anteriores, es decir, se mantiene la búsqueda exhaustiva mediante máscara explicada

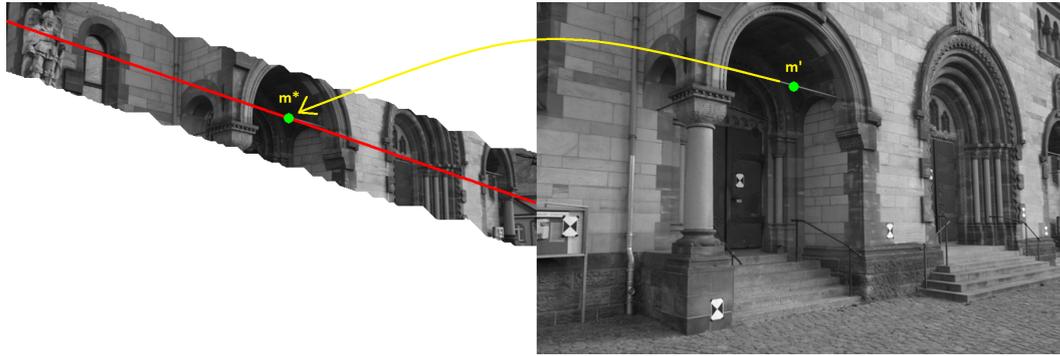


Figura 3.6: *Proyección y retroproyección con el uso de la máscara sobre la línea epipolar.*

anteriormente en 3.4.2.2 y, además, se añade el proceso de retroproyección que definíamos en 3.4.2.1. Es decir, la proyección es igual a la que se muestra en la Figura 3.5 pero, como podemos ver en la Figura 3.6, una vez obtenido el punto m' en B se inicia la fase de retroproyección con el cálculo de la línea epipolar sobre la imagen A y la máscara de búsqueda asociada, ampliando así el área de búsqueda de correspondencias no solo en proyección, sino también en retroproyección. En esta técnica también se permite el procedimiento de doble proyección descrito al final de la sección 3.4.2.1.

3.5. Reestimación de la matriz fundamental

La combinación de las técnicas explicadas en la sección anterior genera un nuevo conjunto de correspondencias con las que se reestima la matriz fundamental. De nuevo, esta reestimación se apoyará en el algoritmo RANSAC (explicado brevemente en la Sección 2.1.3) para estimar una nueva matriz fundamental a partir del nuevo conjunto de correspondencias.

En el Capítulo 5 se evaluará el error producido por la estimación inicial de la matriz fundamental así como por las reestimaciones alcanzadas después de cada una de las fases de refinamiento. Permitiendo así comparar las diferentes aproximaciones para el refinado y verificando la bondad de sus resultados.

3.6. Rotación y Traslación

Obtenida la matriz fundamental, es decir, la matriz que modela la geometría epipolar existente entre las imágenes analizadas, podemos extraer las matrices de proyección, P y P' , de las cámaras que han capturado dichas imágenes. Esto se hace

como se ha explicado en el apartado “Posición relativa de una cámara respecto de la otra” de la Sección 2.1.3.

3.7. Limitaciones

Cuando el desplazamiento entre las imágenes es muy grande los métodos analizados para la extracción automática de correspondencias no son, generalmente, capaces de encontrar un número suficiente de correspondencias correctas con las que obtener una buena estimación inicial de la matriz fundamental. Por eso, y con el fin de poder analizar las estrategias desarrolladas durante el trabajo fin de grado con independencia de la etapa de estimación inicial, se propone diseñar una interfaz de usuario que permita establecer manualmente las correspondencias con las que se realizará la primera estimación de la matriz fundamental. Esta interfaz se explica paso a paso y con todas sus funcionalidades en el siguiente capítulo.

Capítulo 4

Interfaz gráfica de usuario (GUI)

Cuando los desplazamientos entre las imágenes A y B son demasiado grandes los métodos explicados en las Secciones 3.2 y 3.3 no generan buenas estimaciones iniciales de la matriz fundamental ya que las correspondencias encontradas entre ambas imágenes no son, en general, correctas. Para estos casos, y con el fin de probar el funcionamiento de las técnicas explicadas en la Sección 3.4 con independencia de la estimación inicial, hemos implementado una interfaz que permita al usuario seleccionar las correspondencias que considere correctas y oportunas para generar la estimación inicial de la matriz fundamental. En este capítulo describiremos, paso a paso, el funcionamiento de dicha interfaz. El aspecto general de la interfaz diseñada se muestra en la Figura 4.1. Sirva este capítulo como una guía de usuario de la interfaz implementada.

En primer lugar será necesario abrir las dos imágenes, esto se hace pulsando sobre los botones '*Seleccionar Imagen A*' y '*Seleccionar Imagen B*' que podemos encontrar señalados con rojo en la parte superior de la interfaz en la Figura 4.2.

En esta misma Figura 4.2, vemos otros dos botones resaltados con el color verde, con ellos calcularemos la primera estimación de F . Existen para ello dos posibilidades:

1. Pulsar el botón '*Comenzar Estimación Inicial de F*', donde se nos pedirá seleccionar sobre la imagen A cinco puntos de test. Una vez hecho, pasará a la imagen B dónde nos pedirá que establezcamos las correspondencias con los puntos seleccionados anteriormente en A (en el mismo orden de introducción). Estas correspondencias son guardadas en un archivo *.mat* con el nombre "*CorrespondenciasTEST_carpetaImágenes_nombreIA_nombreIB.mat*" y serán utilizadas

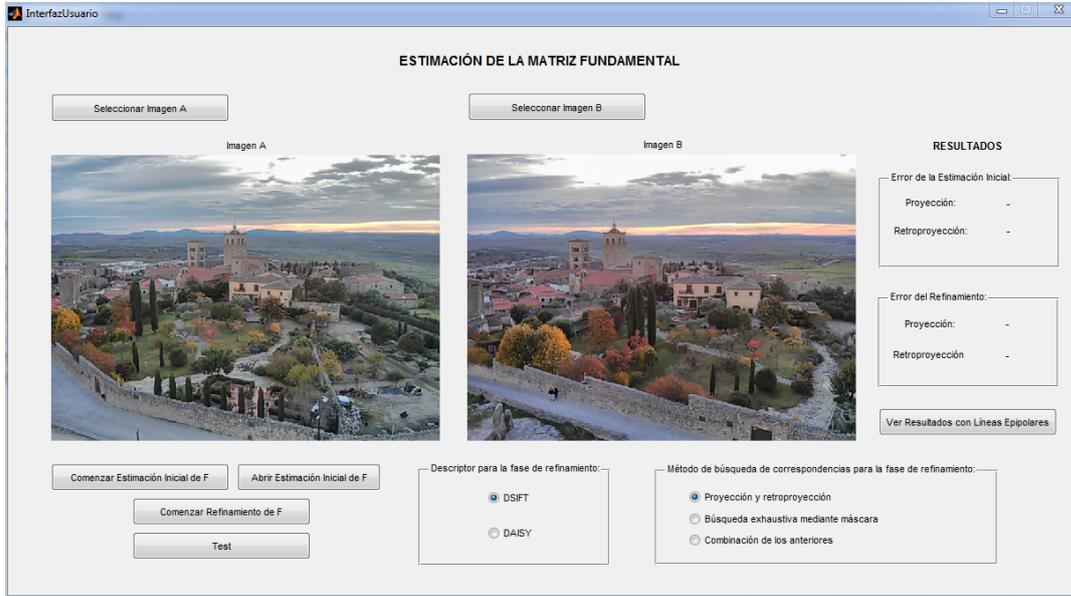


Figura 4.1: *Interfaz de usuario.*

posteriormente para el cálculo de los errores producidos por la matriz fundamental. Después de esto, la interfaz nos volverá a solicitar otros 15 puntos en A y sus respectivas correspondientes sobre B . Estas serán las correspondencias que se introducirán en el algoritmo RANSAC para generar la estimación inicial de F_0 . Una vez finalizado este el proceso, se calcularán tanto los descriptores DAISY como los descriptores DSIFT de las imágenes completas y se almacenará todo ello en un archivo *.mat* con el nombre “*CorrespondenciasF0_[carpetaImágenes]_[nombreIA]_[nombreIB].mat*”.

2. Pulsar sobre el botón ‘*Abrir Estimación Inicial de F*’. Esta opción sólo es posible si alguna vez se ha hecho la estimación inicial pulsando el botón ‘*Comenzar Estimación Inicial de F*’ para el dataset analizado, es decir, sólo si tenemos guardados los archivos *.mat* anteriores. Pulsando sobre este botón únicamente cargaremos los parámetros de dichos archivos para que la interfaz trabaje con ellos sin necesidad de volver a seleccionar las correspondencias manualmente.

Una vez obtenida la estimación inicial de la matriz fundamental podemos probar las diferentes combinaciones de las técnicas desarrolladas, para ello, únicamente es necesario seleccionar una de las opciones del panel denominado ‘*Descripción para la fase de refinamiento*’, es decir, DSIFT o DAISY; y una de las opciones del panel ‘*Método de búsqueda de correspondencias para la fase de refinamiento*’. Una vez escogidos

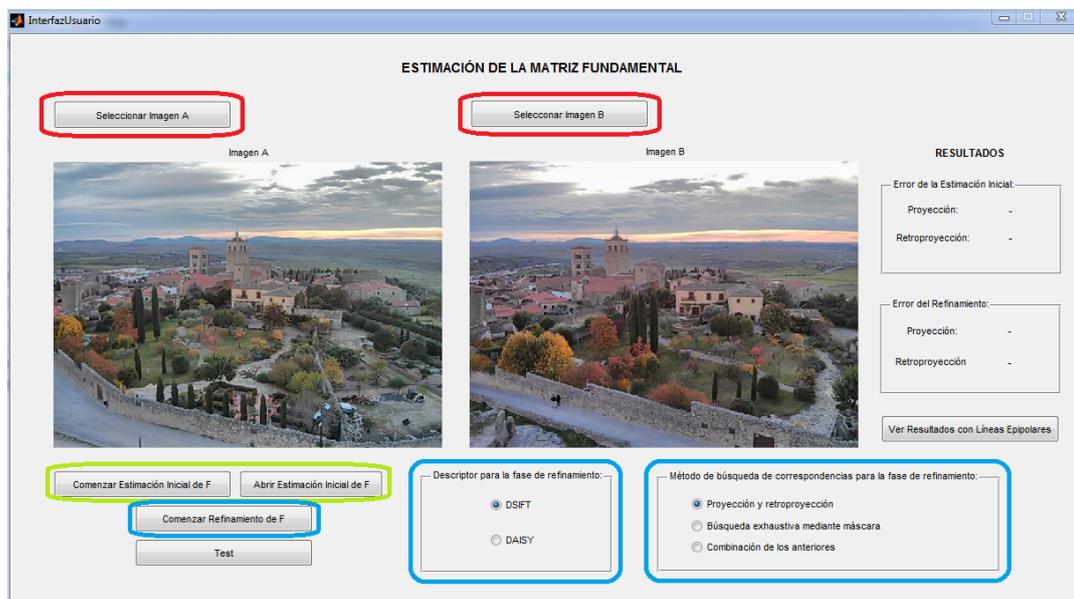


Figura 4.2: *Interfaz de usuario: botones de selección de imágenes (rojo), fase de estimación inicial (verde) y fase de refinamiento (azul)*

estos parámetros sólo tendríamos que pulsar sobre el botón '*Comenzar Refinamiento de F*' para iniciar la fase de reestimación. Tanto los paneles, como el botón nombrado anteriormente podemos encontrarlos resaltados en color azul en la Figura 4.2.

Por último, se procedería al análisis de los resultados pulsando sobre el botón '*Test*' resaltado en color rojo en la Figura 4.3. Estos resultados se basan en la evaluación de las correspondencias de test sobre la matriz fundamental inicial y sobre la matriz fundamental final, es decir, tras el refinamiento. La evaluación se hace calculando la media de la distancia entre los puntos de test y las líneas epipolares de las dos matrices fundamentales, se explicará en detalle en la Sección 5.4.3.

Estos resultados obtenidos pueden observarse sobre los paneles resaltados en amarillo en la derecha de la Figura 4.3. Además tenemos otra forma de ver los resultados sobre las imágenes (ver de nuevo la Figura 4.3) mediante las líneas epipolares que producen sobre B ambas matrices fundamentales (inicial y refinada), para ello únicamente tenemos que pulsar sobre el botón '*Ver Resultados con Líneas Epipolares*'.

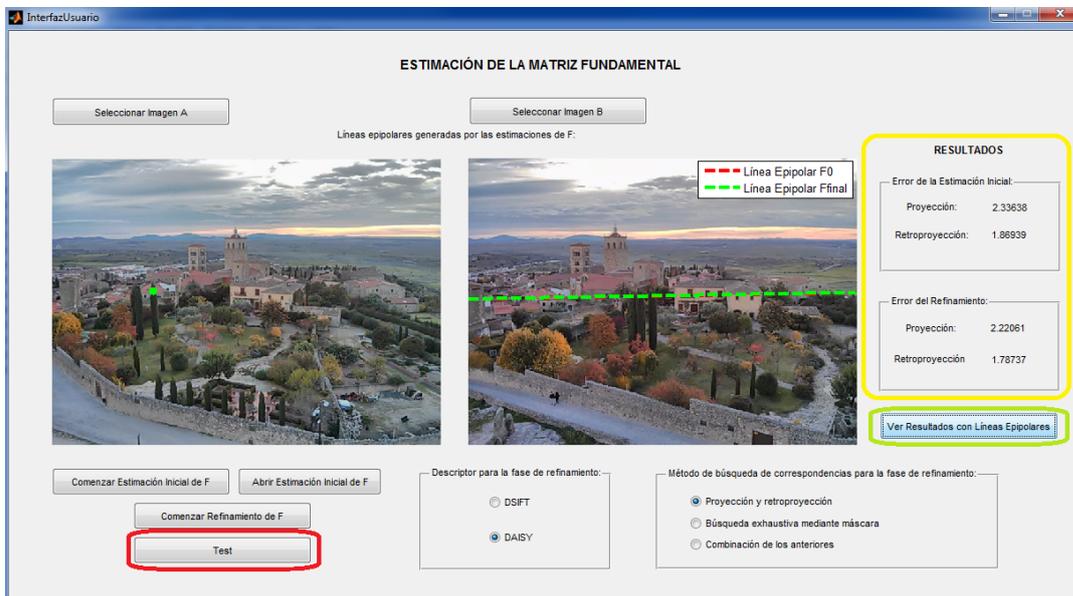


Figura 4.3: *Interfaz de usuario: botón de test para el cálculo de los errores (rojo), paneles dónde se muestran los errores obtenidos (amarillo) y botón para visualizar las líneas epipolares que producen las estimaciones de F (verde).*

Capítulo 5

Resultados experimentales

En este capítulo introduciremos el dataset y el código utilizado para la realización de este trabajo y explicaremos el proceso de evaluación que se ha llevado a cabo así como las medidas de error diseñadas ya sea a través de información extraída del ground-truth o sin ella. Posteriormente analizaremos y compararemos las combinaciones de los diferentes métodos descritos anteriormente en el Capítulo 3.

5.1. Dataset y Código

Principalmente nos hemos centrado en un dataset con ground-truth cuyas imágenes vienen acompañadas de unos archivos de calibración donde podemos encontrar los parámetros intrínsecos y extrínsecos de las cámaras, así como sus matrices de proyección y, además, unos archivos .dmb que contienen las imágenes de profundidad: distancia del punto 3D a la cámara que captura cada una de las vistas. Dentro de este dataset vamos a trabajar durante todo este capítulo sobre las imágenes denominadas Herzjesu que podemos encontrar en el Apéndice B.

Se han utilizado varias librerías de Matlab de código libre. Para la descripción SIFT y DSFIT hemos usado el código que se propone en <http://www.vlfeat.org/>. Por otro lado, el código de búsqueda de correspondencias densas mediante NRDC ha sido extraído de <http://www.cs.huji.ac.il/~yoavhacohen/nrdc/> y, por último la descripción DAISY se ha conseguido gracias al código suministrado en <http://cvlab.epfl.ch/software/daisy>. Hay que destacar que se han mantenido los parámetros por defecto que explican los autores en [12] y [11] tanto para NRDC como para DAISY, respectivamente. Finalmente, para el cálculo de la matriz fundamental a partir de las matrices de proyección y para el procedimiento inverso, es decir, la obtención de las matrices de proyección a partir de la

matriz fundamental se ha utilizado el código suministrado por los autores de [5] en <https://www.robots.ox.ac.uk/~vgg/hzbook/>.

5.2. Procedimiento de evaluación

El dataset utilizado se compone de un total de 7 imágenes. La primera de ellas (*0000.jpg*) la utilizaremos como imagen de referencia durante todos los experimentos y será comparada con el resto de imágenes del dataset. Como se puede ver en el Apéndice B cada una de las imágenes representa una vista más alejada de la imagen de referencia.

El procedimiento de evaluación tiene una fase previa donde se mide el error de la estimación inicial de la matriz fundamental para diferentes parámetros de configuración (número de escalas y /o número de puntos en función del método utilizado). Realizada esta fase, se selecciona la mejor de las estimaciones anteriores, de esta forma se consigue trabajar con la misma estimación inicial para todas las combinaciones posibles de las técnicas diseñadas para el refinamiento, haciendo así más sencilla y fiable la comparación posterior.

Una vez elegida la estimación inicial de la matriz fundamental, se procederá a medir la ganancia o pérdida de error que producen las diferentes combinaciones de las técnicas de refinamiento con respecto al error asociado a la estimación inicial. Para ello hay que obtener el error que deriva de la fase de refinamiento.

5.3. Extracción de la información de *Ground-truth*

Para generar medidas de error correctas nos hemos ayudado de la información real que contiene el dataset, es decir, de su ground-truth. Como ya adelantábamos en la Sección 5.1, cada una de las imágenes viene acompañada de un archivo .dbm con información de profundidad y otro archivo de calibración que contiene, entre otras cosas, los parámetros intrínsecos y extrínsecos de las cámaras que han capturado la imágenes así como sus matrices de proyección.

A partir de las matrices de proyección reales de las dos vistas bajo análisis, extraídas de los archivos de calibración, calculamos la matriz fundamental real (F_{real}) que define la geometría epipolar mediante el procedimiento descrito en [5]. Esta matriz fundamental real nos servirá para medir la bondad de las correspondencias encontradas entre las dos imágenes (Sección 5.4.1).

Por otro lado, tanto las matrices de proyección obtenidas del archivo de calibración como las imágenes de profundidad serán utilizadas para generar una medida de error

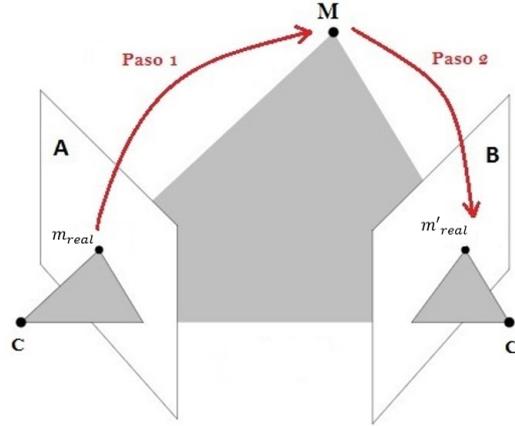


Figura 5.1: Extracción de los mapas de error de F .

que establezca la bondad de las matrices fundamentales estimadas. Como se muestra en la Figura 5.1, partimos de los puntos de la imagen A y, con la información de profundidad y un sencillo cambio de coordenadas mediante los parámetros extrínsecos de la matriz de proyección de la cámara que captura la imagen A , obtenemos los puntos del espacio que fueron capturados por la cámara. El siguiente paso, paso 2 en la Figura 5.1, sería obtener el lugar exacto donde deberían proyectarse estos puntos del espacio sobre la imagen B . De nuevo, esto se consigue con la ayuda de la matriz de proyección de la cámara B . Una vez hecho esto tendríamos la posición real de las proyecciones de todos los puntos de la imagen A sobre la imagen B .

Este proceso puede formularse de la siguiente manera. Sea λ la información de profundidad asociada al punto 2D m_{real} . El punto 3D P^{-1} se obtiene en el paso 1:

$$M = \lambda[P(\cdot, 4)]^{-1}m_{real} - [P(\cdot, 4)]^{-1}P(\cdot, 1, 2, 3) \quad (5.1)$$

, donde $P(\cdot, j)$ es la submatriz resultante de eliminar la columna j y $[P]^{-1}$ es la inversa de la matriz P .

A partir del punto 3D M , su proyección en la otra imagen: m'_{real} se obtiene en el paso 2 como:

$$m'_{real} = P'M \quad (5.2)$$

5.4. Medidas de error

Se han diseñado tres medidas de error diferentes. Dos de ellas se apoyan en el ground-truth: una para evaluar la bondad de las correspondencias y otra para medir la precisión en la estimación de la matriz fundamental. La tercera métrica de error es extraída a partir de unas correspondencias de test elegidas por el usuario mediante la GUI descrita en el Capítulo 4.

5.4.1. Error de las correspondencias con Ground-Truth

A partir de la matriz fundamental real, F_{real} , procederíamos al cálculo del error de las correspondencias mediante la ecuación que define la distancia entre punto y recta. La distancia derivada de dicha ecuación estaría medida en píxeles.

En primer lugar se obtiene el error de proyección partiendo de los puntos en A de las correspondencias a evaluar, generando sus líneas epipolares sobre la imagen B con ayuda de la matriz fundamental real extraída del ground-truth mediante la Ecuación 2.12 y calculando la distancia entre estas rectas y los puntos en B de las correspondencias. Por lo tanto, si partimos de un punto en A (m) cuya correspondencia es el punto m' sobre la imagen B , las operaciones serían las siguientes:

$$l'_{real} = F_{real} \cdot m \quad (5.3)$$

$$d(l'_{real}, m') = \frac{|l'_{real,a} \cdot u' + l'_{real,b} \cdot v' + l'_{real,c}|}{\sqrt{(l'_{real,a})^2 + (l'_{real,b})^2}} \quad (5.4)$$

Donde $l'_{real,a}$, $l'_{real,b}$ y $l'_{real,c}$ son los valores que definen la ecuación general de la línea epipolar sobre la imagen B ; y u' y v' son las coordenadas que definen m' .

De la misma forma, se procedería al cálculo del error de retroproyección. La única diferencia entre ambos es que esta vez partiríamos del punto de la correspondencia que se encuentra sobre la imagen B . Es decir, siendo un punto m' en la imagen B , se genera la línea epipolar correspondiente sobre A mediante $l_{real} = F_{real}^T \cdot m'$, calculando con ello la distancia entre el punto de la correspondencia en A (m) y dicha línea epipolar mediante la Ecuación 5.4.

Esto se haría para todas y cada una de las correspondencias bajo análisis. Por último, y con la finalidad de quedarnos con una única cifra, obtendríamos la media del error entre correspondencias:

$$\varepsilon(m', m | F_{real}) = \frac{d(l'_{real}, m') + d(l_{real}, m)}{2} \quad (5.5)$$

Si se desea obtener una medida global del error para un conjunto de N correspondencias $\Pi = \{m_i \longleftrightarrow m'_i, i \in [1, N]\}$, puede calcularse el error promedio del conjunto:

$$\varepsilon(\Pi|F_{real}) = \frac{1}{N} \sum_i^N \varepsilon(m'_i, m_i|F_{real}) \quad (5.6)$$

5.4.2. Error de la matriz fundamental usando información de Ground-Truth

En este caso se medirá el error de la matriz fundamental partiendo de las posiciones reales de las proyecciones de todos los puntos de la imagen A sobre la imagen B que fueron calculadas con ayuda del ground-truth (Sección 5.3) siguiendo los pasos de la Figura 5.1. Una vez obtenidas estas proyecciones reales $\Pi_{real} = \{m_{real} \longleftrightarrow m'_{real}\}$ se podría calcular el error de la matriz fundamental obtenida (F_f) entre la proyección real en B (m'_{real}) y las líneas epipolares generadas a partir de F_f de un punto en A m_{real} mediante la ecuación:

$$\varepsilon(F_f|m_{real}) = d(l'_f, m'_{real}) \quad (5.7)$$

Si analizamos el error para todos los píxeles de la imagen, es posible generar un mapa de error para F_f como el que podemos ver en la Figura 5.2.

Por otro lado, para suministrar una medida global de la calidad de la estimación realizada, se propone introducir una medida de error medio para cada estimación de F_f . Para ello, simplemente promediamos los errores obtenidos para el número de correspondencias analizadas (N):

$$\varepsilon(F_f|\Pi_{real}) = \frac{1}{N} \sum_i^N \varepsilon(F_f|m_{real,i}) \quad (5.8)$$

5.4.3. Error de las correspondencias sin utilizar información de Ground-Truth

En el caso en el que se desee analizar el comportamiento de las técnicas diseñadas sobre datasets que no incluyan información de ground-truth, se puede realizar una evaluación de las correspondencias utilizando información suministrada por el usuario.

Para estos casos, el usuario debe introducir manualmente un número determinado de correspondencias correctas, denominadas correspondencias de test ($\Pi_{test} =$

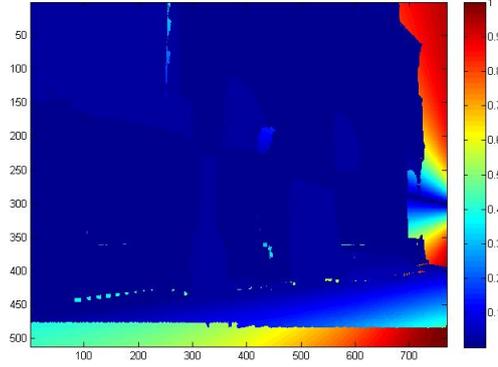


Figura 5.2: Ejemplo de mapa de distribución del error de la matriz fundamental.

$\{m_{test} \longleftrightarrow m'_{test}\}$). A partir de ellas, y con las estimaciones de la matriz fundamental cuya bondad queremos medir (F_f), obtenemos sus líneas epipolares (l'_f , y l_f ,) sobre las imágenes, y calculamos el error de F_f a partir de la información suministrada por el usuario es decir, para un punto de test (m_{test}):

$$\varepsilon(F_f|m_{test}) = d(l'_f, m'_{test}) \quad (5.9)$$

, la interfaz también permite medir el error obtenido en la retroproyección, siguiendo simplemente:

$$\varepsilon(F_f|m'_{test}) = d(l_f, m_{test}) \quad (5.10)$$

Siguiendo el esquema de las secciones anteriores, para calcular una medida de error global en este caso simplemente promediamos el error de todas las correspondencias analizadas, para la proyección:

$$\varepsilon(F_f|\Pi_{test}) = \frac{1}{N} \sum_i^N \varepsilon(F_f|m_{test,i}) \quad (5.11)$$

y, para la retroproyección:

$$\varepsilon(F_f|\Pi_{test}^{-1}) = \frac{1}{N} \sum_i^N \varepsilon(F_f|m'_{test,i}) \quad (5.12)$$

Nótese que este error sólo se calculará en el análisis de los resultados obtenidos con la GUI descrita en el Capítulo 4.

5.5. Análisis de resultados

Una vez que hemos explicado cómo medimos la bondad de nuestras estimaciones podemos pasar a analizar y comparar las diferentes técnicas implementadas. Las tablas de datos de los gráficos que se manejarán en esta sección pueden encontrarse en el Apéndice C.

5.5.1. Análisis de F_0 a partir de la búsqueda de correspondencias iniciales

En primer lugar, es necesario aclarar que el algoritmo con el que se realizan las estimaciones de la matriz fundamental, es decir, RANSAC, es de naturaleza no determinista, por lo tanto si introducimos las mismas correspondencias (número y naturaleza) dos veces el resultado obtenido puede ser diferente. Por este motivo y para aislar el funcionamiento del algoritmo de este factor de incertidumbre se ha calculado la estimación inicial de la matriz fundamental, F_0 , una única vez y se ha almacenado en un archivo .mat para su uso en las siguientes secciones.

F_0 basada en búsqueda de correspondencias locales

Como adelantábamos en la Sección 3.2.1, cuando utilizamos el método de búsqueda de correspondencias locales para la estimación inicial de la matriz fundamental, realizamos un barrido de los posibles parámetros de configuración, en este caso únicamente el número de puntos (N) y el número de escalas (nsc) debido a que la forma del detector se ha fijado a detector *con signo*. Con cada una de las configuraciones se obtiene una estimación inicial de la matriz fundamental mediante RANSAC para la cual se calcula y almacena su error. Posteriormente, utilizaremos aquella configuración cuya matriz fundamental presente menor error medio entre todas las correspondencias clasificadas por RANSAC como *inliers* (N_{RANSAC}), es decir, a las utilizadas para la estimación de F_0 . Introduciendo estos parámetros en la Ecuación 5.8 se obtiene:

$$\varepsilon(F_0|\Pi_{real}) = \frac{1}{N_{RANSAC}} \sum_i^{N_{RANSAC}} \varepsilon(F_0|m_{real,i}) \quad (5.13)$$

En la Tabla 5.1 se muestra un ejemplo de este proceso de estimación inicial barriendo ambos parámetros para un par de imágenes, los resultados obtenidos, siguiendo el mismo proceso, para el resto de imágenes analizadas, se incluyen en el Apéndice C.

En la Tabla 5.1, se representa el número de puntos en las columnas y el número de escalas en las filas. La matriz fundamental con menor error la encontramos resaltada

nsc \ N	100	250	500	1000	1500	2000	2500	3000
10	229,936	30,5991	48,9789	17,2188	10,6087	35,1625	16,6409	48,6875
20	24,7542	58,6726	31,2702	35,3093	41,2243	14,2396	26,6471	40,8808
30	17,3044	48,0403	11,9336	13,6616	40,2902	26,5676	22,4297	17,3805
40	44,5508	31,1913	42,9526	6,0631	52,0788	39,2011	34,2435	28,0786
50	158,7535	18,823	40,5948	6,9412	3,9369	8,0492	15,3282	56,2886
60	177,9003	22,6988	52,3426	50,7884	19,9277	7,9623	47,7601	18,1459
70	18,6891	89,9998	43,227	7,766	13,5108	17,7527	29,5984	21,7117
80	11,8534	25,0639	38,7076	5,1351	15,2013	13,6477	48,571	7,7303
90	35,4087	19,3073	31,1712	7,5946	13,94	6,6137	17,3571	59,3063
100	96,1985	22,9882	34,1639	17,6631	44,8221	21,0477	23,563	46,5978

Tabla 5.1: *Búsqueda automática de correspondencias locales y matriz fundamental asociada: barrido de los parámetros de configuración y error medio $\varepsilon(F_0)$ resultante para cada configuración analizada.*

en rojo y se corresponde con la configuración $nsc = 50$ y $N = 1500$ (para este caso particular resulta $N_{RANSAC} = 475$ como se puede ver en la Tabla 5.2).

Realizamos este proceso para cada par de imágenes del dataset Herzjesu, manteniendo la imagen *0000* como la de referencia, y los resultados obtenidos los podemos ver en la Tabla 5.2. En esta tabla también se ha incluido el error medio entre las correspondencias que RANSAC clasifica como *inliers*, es decir, aquellas que han sido utilizadas para la estimación de F_0 . Este error (en la tercera fila de la Tabla 5.2), se mide respecto a la matriz fundamental real, es decir, es el promedio resultante de medir el error definido en la Ecuación 5.6 para las N_{RANSAC} correspondencias (segunda fila de la Tabla 5.2). Introduciendo estos parámetros en la Ecuación 5.6, obtenemos:

$$\varepsilon(\Pi|F_{real}) = \frac{1}{N_{RANSAC}} \sum_i^{N_{RANSAC}} \varepsilon(m'_i, m_i|F_{real}) \quad (5.14)$$

F_0 basada en búsqueda de correspondencias densas

Mostraremos ahora la tabla de resultados que deriva de la obtención de la matriz fundamental a partir de la búsqueda inicial de correspondencias densas mediante de NRDC. Recordamos que en este caso se ha hecho una preselección de $N = 3000$ correspondencias sobre el conjunto de aquellas que tenían máxima confianza con el fin de reducir el coste computacional. Es decir, para este caso no se realiza un barrido de los

	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
Nº escalas	50	20	10	90	60	90	70
Correspondencias	475	173	53	165	51	30	77
Error Correspondencias F0	9,8742947	35,213601	15,922612	252,12011	271,05737	220,28594	250,28603
ErrorF0	3,936854	41,834136	128,22428	1118,0277	147,10813	121,42739	114,10959

Tabla 5.2: *Mejor estimación de F_0 para cada par de imágenes a partir de búsqueda automática de correspondencias locales. Incluyendo: la escala óptima de extracción (nsc), el número de inliers devuelto por RANSAC (N_{RANSAC}), el error medio asociado para la matriz fundamental estimada y para las correspondencias utilizadas para su estimación.*

parámetros. Los resultados obtenidos después de este proceso de preselección se muestran sobre la Tabla 5.3. Nótese, que el número final de correspondencias N_{RANSAC} puede ser menor que el deseado (N), puesto que sólo se contemplan las que el algoritmo RANSAC clasifica como *inliers*.

	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
Correspondencias	2995	2902	2990	2014	2612	3000	2799
Error Correspondencias F0	1,7011652	2,4041211	1,951158	3,0478376	36,405143	2,2880354	30,572105
ErrorF0	8,3742577	6,8186359	19,593322	8021,6338	732,41308	479,68259	345,6729

Tabla 5.3: *Resultados de F_0 a partir de búsqueda de correspondencias densas después de la preselección de correspondencias. Incluyendo: el número de inliers devuelto por RANSAC (primera fila), el error medio asociado para la matriz fundamental estimada (tercera fila) y para las correspondencias utilizadas para su estimación (segunda fila).*

Comparación de las dos formas diferentes de obtener F_0

Podemos comparar los resultados de la matriz fundamental obtenidos mediante el uso de búsqueda de correspondencias locales o densas. A través del gráfico de la Figura 5.3, observamos que, a excepción del análisis de la imagen de referencia con las imágenes *0002* y *0003*, el método SIFT obtiene mejores resultados que la preselección de correspondencias NRDC.

Como veremos a continuación, cuando los errores de la estimación inicial se disparan será complicado obtener una mejora significativa de la matriz fundamental.

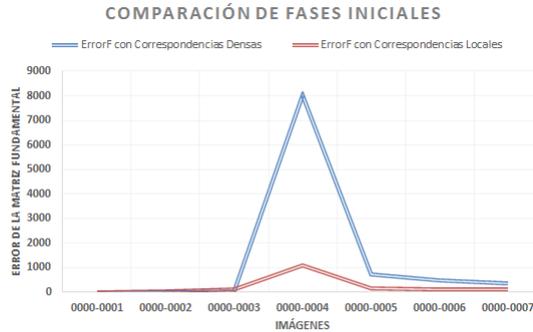


Figura 5.3: Comparativa de los dos métodos de estimación inicial: locales (*SIFT*) y globales (*NRDC*)

5.5.2. Análisis de F tras el refinamiento

En la sección anterior comparábamos los dos métodos iniciales de búsqueda de correspondencias para la estimación de la matriz fundamental, ahora podremos comparar las diferentes técnicas diseñadas para una posible mejora de esa primera estimación de la matriz fundamental.

Resultados derivados de las técnicas de refinamiento desarrolladas para la mejora de las correspondencias.

Error global

Empezaremos comparando las tres técnicas vistas en la Sección 3.4.2 para la mejora de las correspondencias iniciales, para ello fijamos tanto el método elegido para la estimación inicial F_0 , como el tipo de descriptor elegido para la fase de refinamiento. Por ejemplo, en el caso de la Figura 5.4, en la cual el error producido por la estimación inicial queda ocluido bajo el producido por el método de proyección y retroproyección, partimos de la estimación inicial obtenida mediante la búsqueda de correspondencias locales y observamos los errores globales $\varepsilon(F_f)$ que se producen tras la fase de refinamiento mediante las diferentes técnicas diseñadas.

Veamos ahora, con ayuda de la Figura 5.5, los resultados que se producen cuando partimos de la estimación inicial F_0 , derivada de la búsqueda de correspondencias densas mediante *NRDC*, donde una vez más el error de la estimación inicial queda ocluido en la primera gráfica por el de proyección y retroproyección. Si queremos ver

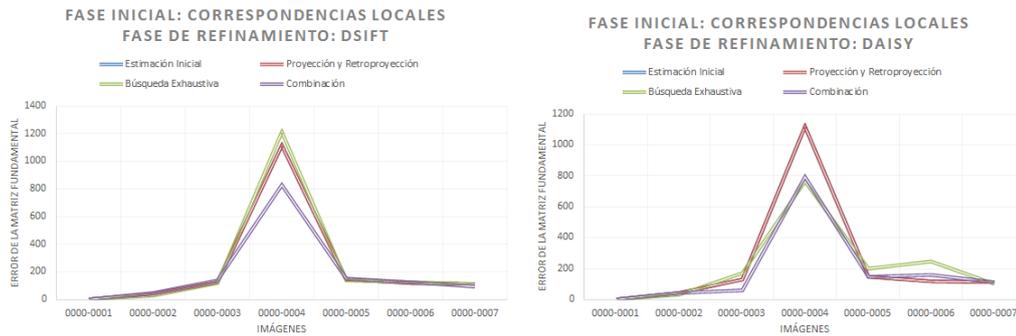


Figura 5.4: Comparación de las técnicas de búsqueda de nuevas correspondencias a partir de la estimación inicial obtenida mediante correspondencias locales (SIFT).

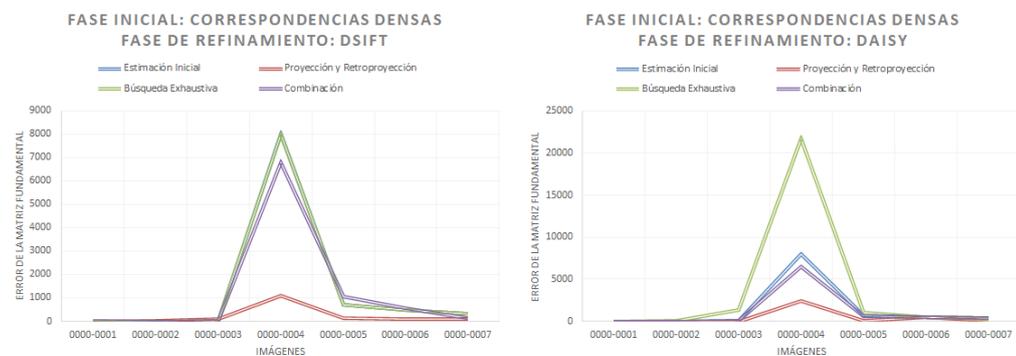


Figura 5.5: Comparación de las técnicas de búsqueda de nuevas correspondencias a partir de la estimación inicial mediante correspondencias densas (NRDC).

los resultados en tablas, debemos ir al Apéndice C dónde numéricamente serán más claros.

Distribución del error en la imagen

Los números que vemos en los gráficos anteriores y en sus correspondientes tablas (Apéndice C) proceden de la media extraída del mapa de error de la matriz fundamental.

Si el objetivo es analizar las áreas de la imagen para las cuales los métodos analizados fallan (o tienen éxito), la distribución del error producido por las estimaciones de la matriz fundamental puede verse creando los mapas de distribución del error para cada de estos métodos a partir de la Ecuación 5.7 aplicada para todos los píxeles de la imagen referencia.

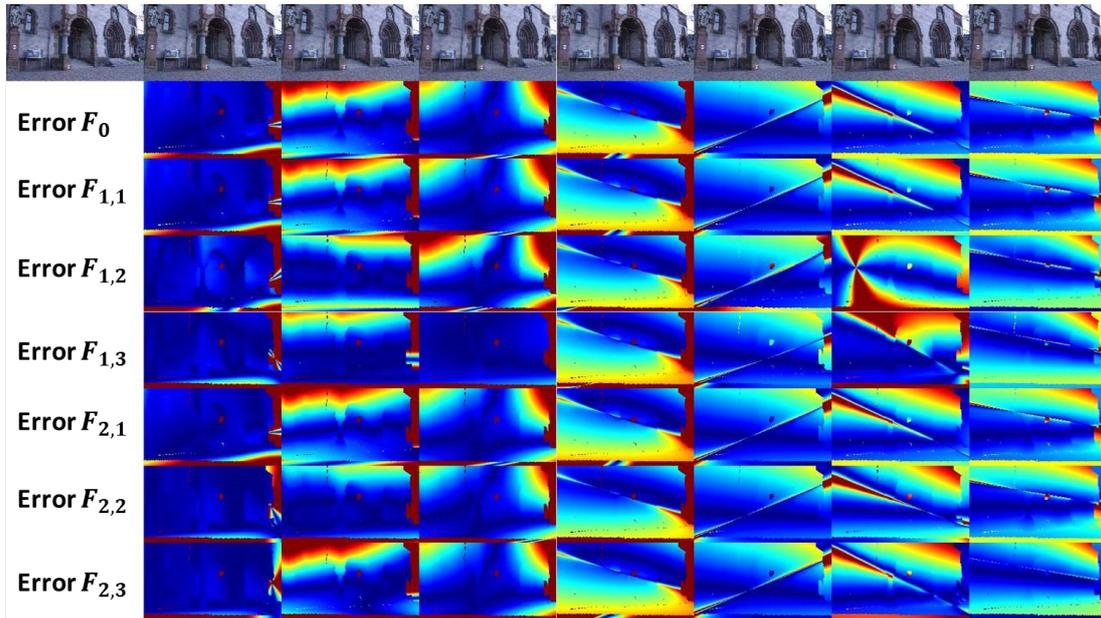


Figura 5.6: Mapas de error utilizando búsqueda de correspondencias locales para la estimación inicial F_0 .

Las Figuras 5.6 y 5.7 muestran los mapas de error obtenidos para todas las posibles combinaciones de la fase de refinamiento. La nomenclatura utilizada para referirnos a estas combinaciones es $F_{i,j}$, donde $i = 1, 2$ (siendo $i = 2$ mejora en la descripción mediante DAISY y $i = 2$ mediante DSIFT) y $j = 1, 2, 3$ (siendo $j = 1$ mejora en la búsqueda de correspondencias mediante proyección y retroproyección, $j = 2$ mediante búsqueda exhaustiva y $j = 3$ mediante combinación de las técnicas anteriores).

Los mapas de error de las Figuras 5.6 y 5.7 han sido obtenidos normalizando respecto al máximo error de la estimación inicial F_0 de la cual parte cada refinamiento y el mapa de color utilizado se encuentra en la Figura 5.8.

Resultados derivados de las técnicas desarrolladas para la mejora en la descripción de las imágenes.

En este apartado se suministra una medida global del error introducido por cada uno de los métodos utilizados para describir los puntos de la imagen, para poder así

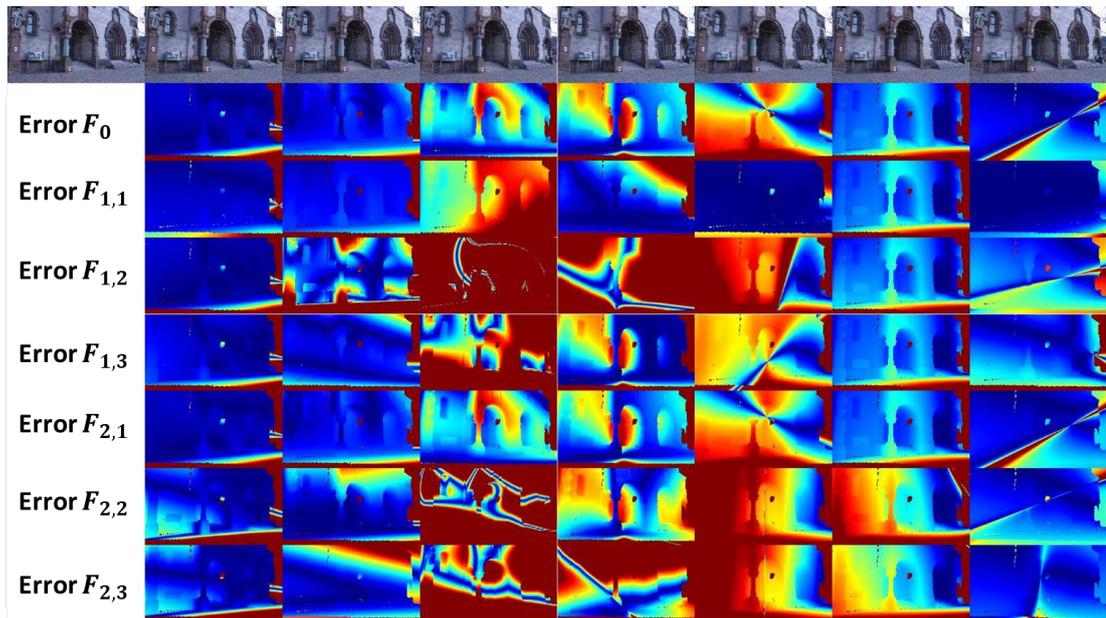


Figura 5.7: Mapas de error utilizando búsqueda de correspondencias densas para la estimación inicial F_0 .

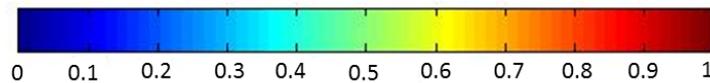


Figura 5.8: Leyenda de color.

evaluar cuál de los métodos de descripción es el más adecuado para cada método de estimación inicial y refinamiento diseñado.

Como podemos ver en las Figura 5.9, 5.10 y 5.11, para poder comparar el funcionamiento aislado de los descriptores DSIFT y DAISY fijaremos tanto el método utilizado para la búsqueda de correspondencias que producen la estimación inicial de la matriz fundamental como el método de búsqueda de correspondencias escogido para la fase de refinamiento pero, en este caso, trabajaremos sobre el error global en las correspondencias (Ecuación 5.6), que es el más afectado por la elección de uno u otro descriptor. De nuevo, las tablas utilizadas para hacer los gráficos se presentan en el Apéndice C.

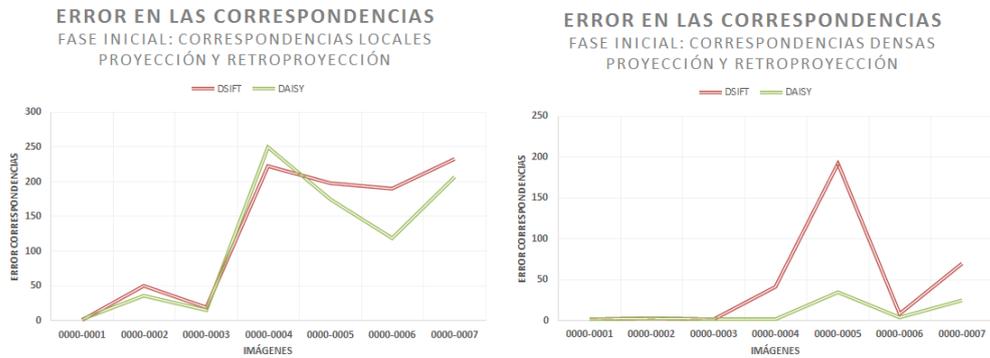


Figura 5.9: Errores en las correspondencias analizando la técnica de Proyección y Retroproyección.

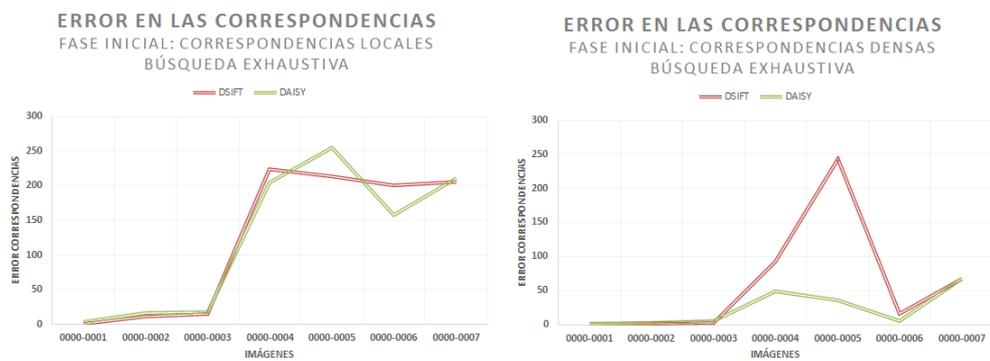


Figura 5.10: Errores en las correspondencias analizando la técnica de Búsqueda Exhaustiva.

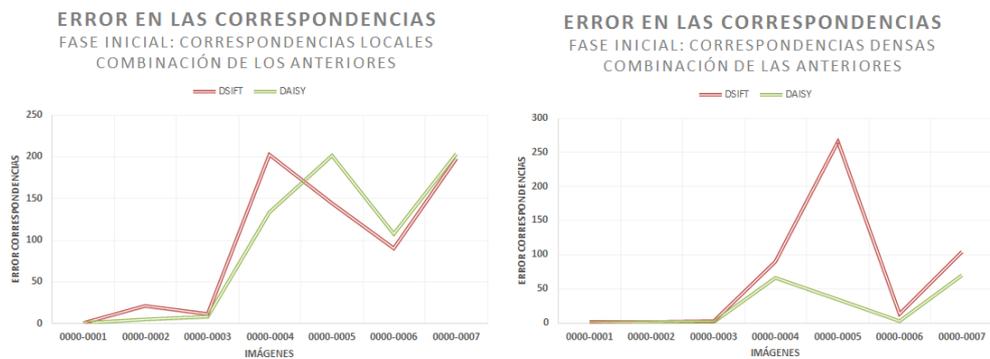


Figura 5.11: Errores en las correspondencias analizando la técnica que combina los métodos de Proyección y Retroproyección y Búsqueda Exhaustiva.

	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
Proyección	1,0294	1,29921	1,56203	1,20283	27,1025	1,80624	26,0858
Retroproyección	0,822875	1,099	1,39677	1,00611	10,8863	1,57436	11,0726

Tabla 5.4: *Estimación inicial de F mediante correspondencias manuales.*

	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
Método	Proyección y retroproyección						
Proyección	1,68828	1,5126	1,52225	1,21965	26,9091	1,79783	25,9808
Retroproyección	1,35317	1,27432	1,38033	1,02371	10,7826	1,54283	11,0586
Método	Búsqueda exhaustiva						
Proyección	1,25154	55,0391	3,48569	26,9506	44,3295	47,2235	56,715
Retroproyección	1,00121	38,7203	3,24043	20,8714	32,5094	330,235	40,2515
Método	Combinación de los anteriores						
Proyección	1,47644	11,0589	2,96393	9,4645	96,8476	101,551	110,789
Retroproyección	1,1767	10,2312	2,65337	10,2366	46,577	60,8809	35,9024

Tabla 5.5: *Refinamiento de correspondencias manuales: Descriptor DSIFT.*

Resultados derivados del uso de la interfaz de usuario.

Como ya se explicaba en el Capítulo 4, hemos implementado una interfaz para que el usuario pueda seleccionar a mano las correspondencias iniciales que se usarán para la estimación inicial de la matriz fundamental.

En primer lugar se seleccionan las imágenes a analizar y $N = 5$ correspondencias de test de las cuales derivarán los resultados. .

Inicialmente analizaremos la bondad de la estimación inicial de la matriz fundamental F_0 , promediando las distancias de las correspondencias de test a las líneas epipolares que ésta produce mediante las Ecuaciones 5.11 (proyección) y 5.12 (retroproyección). Los resultados de este procedimiento se incluyen en la Tabla 5.4.

Por otro lado, se propone medir el error global cometido para cada una de las matrices fundamentales refinadas F_f obtenidas con cada una de las técnicas diseñadas y cada uno de los métodos de descripción evaluados. Para ello aplicamos de nuevo las Ecuaciones 5.11 (proyección) y 5.12 (retroproyección) para cada matriz fundamental. Los resultados obtenidos se incluyen en la Tabla 5.6.

	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
Método	Proyección y retroproyección						
Proyección	1,20984	1,26531	1,56146	1,53925	26,9871	1,61134	26,5286
Retroproyección	0,964344	1,07075	1,40056	1,28872	10,7166	1,45141	11,2816
Método	Búsqueda exhaustiva						
Proyección	0,864439	2,53286	6,27475	4,28698	56,5093	43,8307	91,955
Retroproyección	0,708706	2,0884	5,35096	2,97687	25,7754	27,871	64,733
Método	Combinación de los anteriores						
Proyección	2,31777	3,66546	1,73789	5,53247	13,1124	33,1841	74,3922
Retroproyección	1,91171	2,97076	1,56391	4,84984	7,59112	29,0595	51,5236

Tabla 5.6: Refinamiento de correspondencias manuales: Descriptor DAISY.

5.6. Estimación de la posición de una cámara respecto de otra

Finalmente, evaluaremos la capacidad de estimar la posición de una cámara respecto de la otra a partir de las matrices fundamentales obtenidas con cada método descrito. Para ello, realizaremos el proceso de obtención de las matrices de rotación y traslación descritos en la Sección 2.1.3.

Las Figuras 5.12 y 5.13 comparan el error cometido en cada estimación de la traslación y las Figuras 5.14 y 5.15 el error de las estimaciones de las matrices de rotación para cada método analizado. Dichas tablas siguen la nomenclatura $F_{i,j}$ definida en el análisis de los mapas de distribución del error de F_f .

El error se mide respecto al vector de traslación \vec{t}_{real} y la matrices de rotación R_{real} entre cada par de imágenes. En particular, sea F_f la estimación de la matriz fundamental analizada y \vec{t}_f y R_f ses vectores de traslación y matriz de rotación asociados, el error cometido en la estimación de la traslación se mide mediante la norma L2 como:

$$\varepsilon(\vec{t}_{real}, \vec{t}_f) = \left\| \vec{t}_{real} - \vec{t}_f \right\|_2 \quad (5.15)$$

y el error en la estimación de la matriz de rotación, por medio de la norma de Frobenius:

$$\varepsilon(R_{real}, R_f) = \|R_{real} - R_f\|_{fro} \quad (5.16)$$

Error de los vectores de traslación para una fase de estimación inicial mediante correspondencias densas

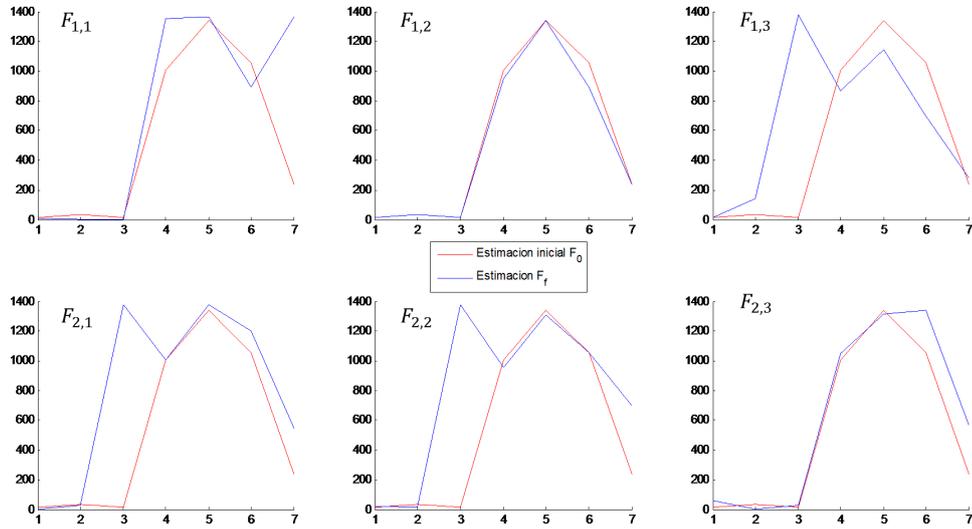


Figura 5.12: Error de los vectores de traslación para una fase de estimación inicial mediante correspondencias densas.

Error de los vectores de traslación para una fase de estimación inicial mediante correspondencias locales

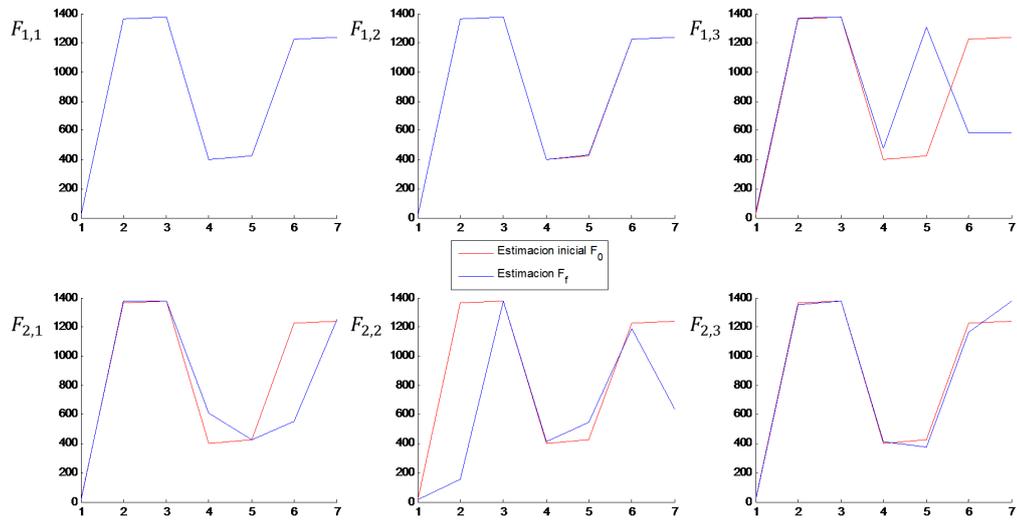


Figura 5.13: Error de los vectores de traslación para una fase de estimación inicial mediante correspondencias locales.

Error de las matrices de rotación para una fase de estimación inicial mediante correspondencias densas

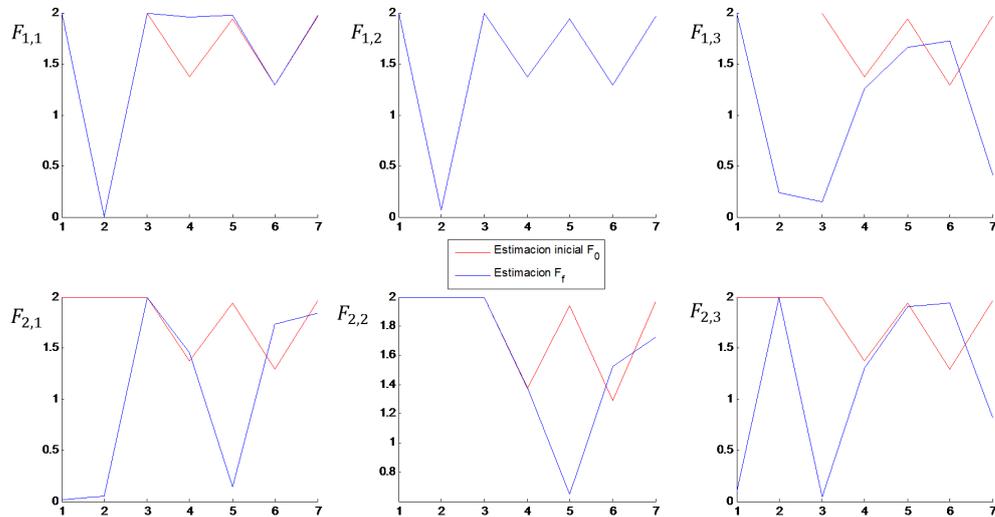


Figura 5.14: Error de los vectores de rotación para una fase de estimación inicial mediante correspondencias densas.

Error de las matrices de rotación para una fase de estimación inicial mediante correspondencias locales

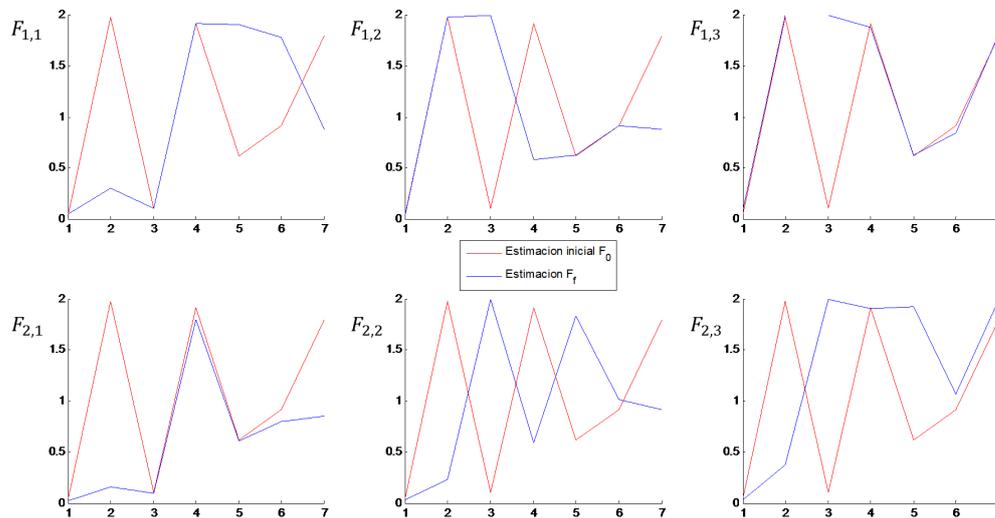


Figura 5.15: Error de los vectores de rotación para una fase de estimación inicial mediante correspondencias locales.

5.7. Evaluación de resultados

A la vista de las tablas y gráficos anteriores podemos evaluar tanto las estimaciones iniciales obtenidas mediante la búsqueda de correspondencias locales o densas, como cada una de las técnicas diseñadas durante el transcurso del trabajo de fin de grado.

En cuanto a las estimaciones iniciales de F , y como podemos ver en las Tablas 5.2, 5.3 y en la Figura 5.3, dichas matrices fundamentales generadas por ambos métodos de búsqueda de correspondencias iniciales producen unos errores mayores conforme la separación entre las vistas aumenta. Esto se debe a que los descriptores elegidos para esta fase no son lo suficientemente robustos a los cambios existentes entre las imágenes cuando la distancia entre las cámaras que las capturan es elevada.

Generalmente, y como podemos ver con ayuda de los gráficos de las Figuras 5.4 (Tablas C.11 y C.12) y 5.5 (Tablas C.13 y C.14), las técnicas de refinamiento que parten de estimaciones iniciales de la matriz fundamental obtenidas mediante correspondencias locales producen mejoras más significativas. En cambio, las derivadas de la búsqueda de correspondencias densas, generan mejoras insignificantes o, incluso, producen peores resultados. Esto puede atribuirse a que el conjunto de correspondencias densas no procede de una fase previa de detección singularidad de puntos de interés sino de la extracción de unos mapas de confianza y, por lo tanto, los puntos de las correspondencias pueden no tener ninguna característica que les diferencie del resto de puntos. Esto queda reflejado en su descripción y, a la hora de buscar correspondencias mediante alguna de las técnicas diseñadas para su mejora, no encontraremos un descriptor que genere una distancia de correspondencia (Ecuación 2.25 de la Sección 2.4.1) que sea, el menos, η veces menor que la obtenida para cualquier otro punto. Lo mismo ocurre cuando el usuario selecciona manualmente las correspondencias que estimarán la matriz fundamental inicial mediante la interfaz de usuario descrita en el Capítulo 4.

Pasemos a hablar ahora de las descripciones que producen DAISY y DSIFT como posibles opciones de la fase de mejora en la descripción de los puntos de las imágenes. Si observamos las Figuras 5.9, 5.10 y 5.11 vemos que, para los mismos escenarios, DAISY generalmente obtiene menores errores en las correspondencias. Por lo que podríamos concluir que es preferible la elección de DAISY frente a DSIFT.

Analicemos los resultados derivados de cada una de las técnicas diseñadas para la fase de mejora de correspondencias, de nuevo, con ayuda de los gráficos de las Figuras 5.4 y 5.5 y sus respectivas tablas en el Apéndice C. Es fácil ver que, generalmente,

la técnica de búsqueda exhaustiva es la que produce mejoras más pequeñas o incluso llega obtener errores mayores a los de la estimación inicial. Esto puede atribuirse a que este método carece de fase de retroproyección. Por otro lado, es lógico intuir el motivo por el cual el método de proyección y retroproyección obtiene resultados parecidos al de la estimación inicial de F , esto puede deberse a que trabaja únicamente buscando correspondencias sobre la línea epipolar inicial, por lo que no nos desviaremos en exceso de dicha aproximación. En cuanto al método que combina proyección y retroproyección con búsqueda exhaustiva se puede concluir que si los descriptores son suficientemente robustos a los cambios existentes entre las imágenes será el método que mayor mejora pueda producir.

No sólo hay que tener en cuenta los resultados que se muestran en los gráficos 5.4 y 5.5, puesto que han sido generados calculando la media de error de F_f mediante la Ecuación 5.8, sino que también es necesario entender y comparar la distribución de dicho error mediante las Figuras 5.6 y 5.7. Estas Figuras apoyan las conclusiones extraídas en el apartado anterior donde, por ejemplo, podemos ver como la búsqueda exhaustiva produce, además de peores medias de error, peores distribuciones.

Por último, se han evaluado los errores de las matrices de rotación y de los vectores de traslación extraídas de cada una de las estimaciones de F obtenidas. De los gráficos de la Figura 5.12, donde la estimación inicial se realiza mediante búsqueda de correspondencias densas, se puede concluir que las estimaciones iniciales y las mejoras desarrolladas con respecto a dichas estimaciones producen vectores de traslación correctos cuando el desplazamiento entre imágenes no es muy grande. En cambio, en la Figura 5.13, donde la estimación inicial se realiza mediante búsqueda de correspondencias locales, esto no se cumple (salvo para el primer par de imágenes). Si analizamos las matrices de rotación mediante las Figuras 5.14 y 5.15 vemos, que para ambos métodos de búsqueda de correspondencias iniciales, las diferentes combinaciones de técnicas de refinamiento normalmente generan mejoras o mantienen las iniciales. A modo de conclusión de este último párrafo, establecemos que la búsqueda de correspondencias densas para las estimaciones iniciales de F produce mejores resultados en cuanto a la obtención de los parámetros extrínsecos de las cámaras, que cuando se calculan las estimaciones iniciales de F mediante búsqueda de correspondencias locales.

Las conclusiones generales que pueden ser extraídas de este análisis de resultados son las siguientes:

- SIFT a pesar de tener un error global menor o similar el de NRDC produce peores estimaciones de F_0 atendiendo a la distribución del error.

- DAISY parece ser mejor que DSIFT en la descripción de puntos de interés.
- Las técnicas para el refinamiento mejoran el error global de las estimaciones iniciales de F que parten de correspondencias locales en mayor medida que aquellas que parten de correspondencias densas (obtenidas mediante NRDC). Esto es lógico ya que las estimaciones dependientes de correspondencias locales son algo peores y por tanto tienen un margen mayor de mejora.
- Ni las estimaciones iniciales, ni las fases de refinamiento tienen éxito en situaciones de grandes desplazamientos.
- La GUI esta diseñada para hacer frente a situaciones de grandes desplazamiento donde, además, las técnicas de refinamiento mejoran sensiblemente la estimación inicial.

Capítulo 6

Conclusiones y trabajo futuro.

6.1. Conclusiones.

Este trabajo tenía como objetivo la obtención de técnicas que consiguieran una estimación de la matriz fundamental que englobara las restricciones establecidas por la geometría epipolar existente entre dos vistas de una misma escena. A partir de dicha estimación se podría obtener la posición de una de las cámaras con respecto de la otra, consiguiendo con esto una calibración externa de las cámaras de forma automática, es decir, sin necesidad de patrones.

Para cumplir este objetivo se ha realizado un estudio del estado del arte de diferentes formas de extracción de correspondencias entre dos imágenes: detectores sobre el Scale-Space, descriptores SIFT-DSIFT y DAISY y la obtención directa de correspondencias mediante NRDC. También se ha hecho un estudio previo de conceptos básicos necesarios para comprender el proceso de captura de imágenes así como la geometría epipolar que relaciona dos vistas de una misma escena.

A partir de la detección de puntos sobre el Scale-Space y la descripción de los mismos mediante SIFT o directamente mediante NRDC, se han extraído correspondencias entre dos imágenes con el fin de generar, mediante RANSAC, una estimación inicial de la matriz fundamental.

Posteriormente, se ha implementado una segunda fase que contiene diferentes técnicas que tienen como objetivo intentar mejorar dicha estimación inicial. Esta segunda fase se compone de un primer proceso de descripción de las imágenes completas a través de DSIFT o DAISY y un segundo proceso que trata de mejorar las correspondencias de la primera estimación a partir de la geometría epipolar derivada de la matriz fundamental inicial. Este segundo proceso se compone de tres técnicas denominadas: proyección y retroproyección, búsqueda exhaustiva y combinación de ambas.

Se han calculado los errores producidos por las dos fases en las que se divide el trabajo, se han comparado las diferentes técnicas y, además, se ha implementado una interfaz de usuario para los casos en los que no se consigue generar una buena estimación inicial de la matriz fundamental. Dicho demostrador solicitará que el usuario introduzca cierto número de correspondencias manualmente para poder realizar un cálculo correcto de la matriz fundamental.

Aunque el rendimiento de las técnicas de mejora desarrolladas depende de la estimación inicial de la matriz fundamental extraída mediante técnicas conocidas y no deriva en mejoras significativas de dicha estimación inicial, podemos decir que se han cumplido los objetivos establecidos al comienzo del proyecto. Los resultados no son los esperados debido a que los descriptores utilizados no son lo suficientemente robustos a cambios de vista tan significativos como los tratados en este trabajo.

Ahora, conociendo las limitaciones encontradas durante el desarrollo del trabajo de fin de grado, podemos establecer diversas líneas de trabajo futuro que podrían establecer mejoras mayores de las estimaciones iniciales de la matriz fundamental

6.2. Trabajo futuro.

A partir de los resultados obtenidos y del estudio previo hecho antes del comienzo del trabajo, podemos establecer diversas líneas de trabajo futuro.

En primer lugar, en las tablas del Apéndice C encontramos numerosas mejoras en los resultados sobre las estimaciones iniciales de F que, a veces, y debido a que los errores de esas primeras estimaciones son muy grandes, se quedan cortos para la obtención de un resultado decente. Por lo que una posible línea de trabajo futuro sería simplemente partir de un mayor número de correspondencias para la obtención de la matriz fundamental inicial y realizar el proceso de refinamiento de forma iterativa hasta que se dejen de obtener mejoras.

Por otro lado, sería interesante probar las técnicas de mejora implementadas con diferentes detectores y descriptores de puntos de interés para la búsqueda de correspondencias iniciales, puesto que, cuánto mejores sean dichas correspondencias, mayores mejoras podrán conseguirse a la hora del refinamiento. De la misma forma, podríamos probar con diferentes algoritmos para la estimación robusta de F a partir de correspondencias y quedarnos con aquel que produzca unas matrices fundamentales con menor error. El objetivo de estas dos líneas de trabajo futuro sería la obtención de una matriz fundamental inicial con menor error posible.

Por último, a la hora de mejorar las correspondencias iniciales mediante los tres métodos desarrollados durante el transcurso del trabajo (proyección y retroproyec-

ción, búsqueda exhaustiva y combinación de las anteriores), podríamos establecer una búsqueda de correspondencias que no sólo tuviera en cuenta la descripción de los puntos bajo análisis sino también la de sus vecinos.

Bibliografía

- [1] Q.-T. Luong and O. Faugeras, “Camera calibration, scene motion and structure recovery from point correspondences and fundamental matrices,”
- [2] O. Faugeras, *Three-dimensional computer vision: a geometric viewpoint*. MIT press, 1993.
- [3] O. Faugeras and Q.-T. Luong, *The geometry of multiple images: the laws that govern the formation of multiple images of a scene and some of their applications*. MIT press, 2004.
- [4] J. Garcia, “Autocalibración y sincronización de múltiples cámaras ptz,” *PFC in UAM*, 2007.
- [5] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [6] Q.-T. Luong and O. D. Faugeras, “The fundamental matrix: Theory, algorithms, and stability analysis,” *International journal of computer vision*, vol. 17, no. 1, pp. 43–75, 1996.
- [7] X. Armangué and J. Salvi, “Overall view regarding fundamental matrix estimation,” *Image and vision computing*, vol. 21, no. 2, pp. 205–220, 2003.
- [8] F. Espuny and P. Monasse, “Singular vector methods for fundamental matrix computation,” in *Image and Video Technology*, pp. 290–301, Springer, 2014.
- [9] T. Lindeberg, *Scale-space theory in computer vision*. Springer Science & Business Media, 1993.
- [10] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [11] E. Tola, V. Lepetit, and P. Fua, “Daisy: An efficient dense descriptor applied to wide-baseline stereo,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32, no. 5, pp. 815–830, 2010.
- [12] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski, “Non-rigid dense correspondence with applications for image enhancement,” in *ACM Transactions on Graphics (TOG)*, vol. 30, p. 70, ACM, 2011.

- [13] T. Lindeberg, *Image matching using generalized scale-space interest points*. Springer, 2013.

Apéndice A

Transformaciones para el cálculo de P

Conversión de las coordenadas de la cámara (x_c, y_c, z_c) a las del plano de la imagen (u, v) :

Se trata de un cambio de 3 a 2 dimensiones. Ayudándonos de la geometría proyectiva podemos relacionar estas coordenadas fácilmente mediante triangulación, obteniendo:

$$\frac{f}{z_c} = \frac{u}{x_c} = \frac{v}{y_c} \quad (\text{A.1})$$

O de forma equivalente:

$$u = f \cdot \frac{x_c}{z_c} \quad v = f \cdot \frac{y_c}{z_c} \quad (\text{A.2})$$

Podemos escribir esto matricialmente como sigue:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{f}{z_c} \begin{bmatrix} x_c \\ y_c \end{bmatrix} \quad (\text{A.3})$$

Por último, pasamos a coordenadas homogéneas:

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad (\text{A.4})$$

Conversión de las coordenadas del plano de la imagen (u, v) al sistema de coordenadas normalizadas de la imagen (u_n, v_n) :

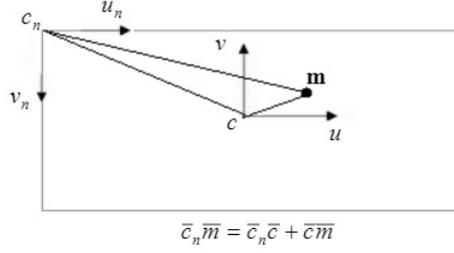


Figura A.1: Cambio del sistema de coordenadas de la imagen al sistema de coordenadas normalizado [4].

El sistema de coordenadas (u, v) está centrado en el punto c , cuyas coordenadas son (u_0, v_0) , donde intersecan el eje óptico con el plano de la imagen y tiene unidades iguales sobre ambos ejes. En este caso, transformaremos este sistema de coordenadas a otro con centro en c_n , normalmente las dimensiones sobre los nuevos ejes son desiguales. Podemos hacerlo sumando vectorialmente a m la posición de c respecto a c_n como se muestra en la Figura A.1.

Debido a este cambio, las coordenadas dejan de expresarse en milímetros y pasan a expresarse en píxeles, para ello es necesario utilizar los factores de conversión:

$$u_n = d_u \cdot u + u_0 \quad v_n = d_v \cdot v + v_0 \quad (\text{A.5})$$

Podemos expresar este cambio en forma matricial como sigue:

$$\begin{bmatrix} u_n \\ v_n \\ 1 \end{bmatrix} = \begin{bmatrix} d_u & 0 & u_0 \\ 0 & d_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (\text{A.6})$$

Con estos dos primeros pasos hemos construido la matriz de parámetros intrínsecos, que en adelante denominaremos K . De esta forma relacionamos el punto M_c , cuyo sistema de coordenadas es el referido a la cámara (X_c, Y_c, Z_c) , con el punto m_n referido a las coordenadas normalizadas de la imagen. De nuevo expresamos esto en forma matricial y con coordenadas homogéneas:

$$\begin{bmatrix} U_n \\ V_n \\ S_n \end{bmatrix} = \begin{bmatrix} f \cdot d_u & 0 & u_0 & 0 \\ 0 & f \cdot d_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \quad m_n = K \cdot M_c \quad (\text{A.7})$$

Únicamente sería necesario un último paso ya que el sistema de coordenadas de la cámara (X_c, Y_c, Z_c) no suele coincidir con el sistema de coordenadas del mundo (X, Y, Z) , este paso se explica en la Sección 2.1.2.

Apéndice B

Imágenes evaluadas

En este Apéndice únicamente mostraremos las imágenes utilizadas para que el lector pueda hacerse una idea del incremento de la distancia entre vistas comparándolas entre sí.



Figura B.1: *Imágenes a analizar. La primera de las imágenes será la utilizada como referencia, es decir, estará en todos los análisis.*

Apéndice C

Tablas de resultados

Utilizaremos este Apéndice para mostrar todas y cada una de las tablas de resultados obtenidas en el análisis de las imágenes Herzjesu extraídas de <http://cvlab.epfl.ch/software/daisy>.

C.1. Estimación inicial SIFT

Aquí se mostrarán las tablas de resultados que se han obtenido realizando el barrido de los parámetros de configuración (número de puntos y número de escalas).

nsc\N	100	250	500	1000	1500	2000	2500	3000
10	339,7446	262,4781	235,9638	185,8656	219,2359	173,2803	171,4966	143,4509
20	209,0283	366,8021	356,4498	310,4742	136,7635	41,8341	147,1456	193,822
30	205,2361	317,182	144,3061	222,9588	129,8115	79,2189	188,9735	133,1225
40	373,7801	211,955	155,5558	163,6908	88,6836	138,9649	145,7364	102,0772
50	406,0853	233,9801	189,9352	165,4514	60,5369	141,8028	113,9004	163,402
60	347,9856	478,4912	139,7133	155,9618	143,9543	94,5829	164,9062	63,7091
70	265,1683	429,3937	317,0486	158,1321	162,1981	210,7515	228,5491	137,8249
80	270,6531	61,972	191,9455	170,5225	162,0085	134,0027	65,6262	157,5368
90	116,8056	410,9147	488,38	134,8632	145,4075	141,9032	115,9605	90,4701
100	333,4605	208,7806	128,0812	177,1015	172,2522	154,2893	149,1301	159,3583

Tabla C.1: *Búsqueda de correspondencias locales: barrido de los parámetros de configuración entre las imágenes 0000-0002.*

nsc\N	100	250	500	1000	1500	2000	2500	3000
10	1968,94264	908,458764	128,224278	415,663128	463,815741	416,891583	1335,04986	1325,85914
20	2272,56822	1193,51525	1001,9324	678,285637	1136,20001	1067,13417	2063,8345	1110,84313
30	2011,06158	2078,47535	819,122753	618,132068	811,257901	648,428211	305,415215	506,423827
40	974,982235	780,884452	1068,34189	398,870794	1030,42918	944,297757	1028,6659	1073,47057
50	1535,47776	1005,73497	342,063067	1030,46246	1208,37171	995,966469	936,82182	580,611737
60	970,983136	1762,43421	1632,00223	139,301349	946,40944	1276,85935	1215,14533	955,946978
70	2002,19405	376,729812	711,881558	898,230839	1445,7262	519,491078	1318,43732	348,64692
80	1106,4055	2166,96377	1392,89075	1200,37829	1148,49211	1314,41928	1025,84161	1094,68705
90	1194,3473	373,282562	1632,66402	893,386683	306,447262	1211,46338	946,782994	1264,20371
100	372,246908	1264,92259	1611,16437	859,396357	728,796898	1372,75097	1373,19324	1087,60641

Tabla C.2: Búsqueda de correspondencias locales: barrido de los parámetros de configuración entre las imágenes 0000-0003.

nsc\N	100	250	500	1000	1500	2000	2500	3000
10	1566,87901	3107,71541	13280,1068	11759,6352	7001,9696	9051,58573	2062,15843	15447,3072
20	18767,5913	5496,21907	10334,9331	15812,4044	10433,1194	5088,05497	12756,9628	1335,21222
30	22723,6224	9444,60823	16333,0458	25704,1151	19524,0182	7068,97266	9163,53447	7589,4553
40	9719,63775	21342,5015	9235,08333	10285,0307	7427,74991	14549,6687	17755,3388	3140,17278
50	23613,3551	10400,1064	18772,8532	7885,13262	7921,22543	7809,84834	17017,8405	21741,6182
60	19337,4905	20488,6055	15429,3782	22006,6806	24956,739	10504,6749	5337,26277	12776,0268
70	13987,8018	6797,37704	4392,75847	12205,869	2166,7006	13380,0786	13731,682	6876,18729
80	26258,7495	9266,18485	13294,3703	11985,6388	3404,74229	1291,47682	8231,46875	7578,1799
90	5499,25374	9722,13635	18580,631	10166,7849	8863,14309	8438,58921	20532,4297	1118,0277
100	26734,4373	16985,9615	12136,7665	8287,74749	12926,3369	4534,64982	13418,8338	12174,2751

Tabla C.3: Búsqueda de correspondencias locales: barrido de los parámetros de configuración entre las imágenes 0000-0004.

nsc\N	100	250	500	1000	1500	2000	2500	3000
10	384,202619	295,427068	450,093066	469,889291	340,272077	413,399039	598,023832	588,453942
20	805,579771	425,417935	402,677417	432,322891	281,941861	433,041712	438,708355	340,935972
30	1060,37627	371,621482	553,989302	474,403835	651,209603	473,613313	446,649861	536,824646
40	645,410038	509,875187	551,462781	544,715815	646,732123	410,663369	312,718676	404,902272
50	412,664332	415,069528	270,515418	800,984859	419,627833	215,87454	485,41029	754,007422
60	415,49747	481,56842	147,108129	221,223771	204,158941	484,332493	386,636398	556,223663
70	384,25479	448,271875	517,364832	151,092277	479,725068	238,418096	263,960569	232,376276
80	491,439405	249,778632	554,618493	516,521902	513,387634	468,678891	568,115624	809,153489
90	446,052028	393,20568	442,523651	169,799823	454,626803	400,573017	355,339373	216,811617
100	665,170861	298,102091	340,157869	724,459257	306,234	422,263606	460,49314	467,922127

Tabla C.4: Búsqueda de correspondencias locales: barrido de los parámetros de configuración entre las imágenes 0000-0005.

nsc\N	100	250	500	1000	1500	2000	2500	3000
10	389,542515	631,118645	260,829189	294,22238	155,189371	607,361496	297,597352	164,95469
20	706,258675	221,143317	345,756116	339,639606	188,948023	134,538249	237,585064	309,696977
30	348,953398	121,542705	477,758501	319,240584	331,360157	383,842397	266,855564	391,176624
40	260,07043	427,800636	214,591162	254,382245	423,103329	401,737574	351,253309	344,639126
50	283,546587	284,967433	519,788137	380,087039	236,544692	241,715662	267,30143	201,473458
60	284,7739	457,169891	276,433379	137,136076	237,952855	274,995946	200,903723	274,401288
70	307,037869	405,366865	361,330422	221,781825	366,789742	156,28615	416,232292	290,395564
80	400,135336	414,716167	485,101409	387,02539	325,391059	272,300988	292,407691	305,115369
90	235,419377	121,427389	190,553616	281,87223	472,324581	185,68759	301,228439	159,478478
100	182,070596	426,623842	525,381076	251,874612	307,025793	178,568679	156,63717	231,826214

Tabla C.5: Búsqueda de correspondencias locales: barrido de los parámetros de configuración entre las imágenes 0000-0006.

nsc\N	100	250	500	1000	1500	2000	2500	3000
10	145,898139	344,604684	200,037028	506,956677	250,943288	226,924529	275,285995	384,9935
20	302,375587	383,164509	525,892321	210,256689	173,728868	344,988676	344,05003	313,018064
30	394,968141	207,539152	393,780822	276,422745	150,331502	336,431126	257,379264	126,476251
40	206,735066	316,324849	178,07919	505,915581	155,942327	228,547811	294,728506	158,079442
50	312,14644	541,059712	293,528654	463,024358	178,745827	170,695005	229,366592	171,21821
60	401,467203	299,49497	253,588307	361,064674	211,561238	423,453302	325,928482	287,61441
70	374,430291	206,265952	190,543363	114,109587	387,817639	426,304081	148,706818	161,346422
80	311,442294	224,415932	385,143921	332,366403	172,582862	542,7512	172,527642	226,201497
90	385,554902	321,932885	319,870843	467,219271	163,429596	258,717007	216,788235	307,145028
100	152,480701	525,543523	413,035562	167,888747	225,542319	257,150013	287,010788	395,765035

Tabla C.6: Búsqueda de correspondencias locales: barrido de los parámetros de configuración entre las imágenes 0000-0007.

C.2. Refinamiento

En primer lugar, vemos el método de estimación inicial mediante búsqueda de correspondencias locales, a partir del cual se han analizado todas las combinaciones posibles de las técnicas de refinamiento explicadas en la Sección 3.4, obteniendo como resultados las Tablas C.7 y C.8.

Se observan también, en las Tablas C.9 e C.10, los resultados obtenidos sobre la estimación inicial realizada a partir de búsqueda de correspondencias densas con NRDC.

	Imágenes	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
Proyección y Retroproyección	Nº correspondencias	463	141	42	131	25	19	46
	Error Correspondencias	1,5623756	36,472364	14,669791	249,90195	174,78349	118,41051	206,16055
	Error F	3,9074429	41,77396	127,82388	1118,1701	147,15108	120,95463	114,21799
Búsqueda Exhaustiva	Nº correspondencias	412	123	42	66	24	12	44
	Error Correspondencias	3,1183583	15,919561	17,724618	204,36821	255,20538	157,12322	211,0214
	Error F	3,1998791	31,83911	168,66222	766,50165	198,46037	245,192	103,5946
Combinación de las anteriores	Nº correspondencias	424	110	29	41	18	11	28
	Error Correspondencias	1,1206097	5,0296068	7,7432034	133,33346	201,79092	107,08238	203,46795
	Error F	3,4258877	43,442155	60,593494	793,18146	148,31359	159,06066	111,18371

Tabla C.7: Resultados completos del refinamiento de la matriz fundamental extraída mediante correspondencias iniciales locales con el descriptor DAISY.

	Imágenes	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
Proyección y Retroproyección	Nº correspondencias	458	149	47	132	28	20	50
	Error Correspondencias	1,216664	49,608271	19,255343	222,24248	197,29567	190,26602	233,21708
	Error F	3,730554	41,796139	127,88727	1117,4913	147,43372	121,33463	114,13307
Búsqueda Exhaustiva	Nº correspondencias	421	120	49	64	22	13	49
	Error Correspondencias	1,3892919	11,056718	15,68309	223,74433	213,48336	200,13016	205,40691
	Error F	2,34964	31,434265	117,69139	1212,9072	139,15962	124,66243	110,1466
Combinación de las anteriores	Nº correspondencias	408	114	40	29	16	10	23
	Error Correspondencias	1,2276691	21,563376	11,559112	202,753	144,35059	90,813454	198,24057
	Error F	1,7787655	50,828099	138,65842	830,10339	149,86062	126,89702	91,002837

Tabla C.8: Resultados completos del refinamiento de la matriz fundamental extraída mediante correspondencias iniciales locales con el descriptor DSIFT.

C.2.1. Error de las matrices fundamentales

Para que al lector le sea más sencillo entender los gráficos del Capítulo 5, adjuntaremos ahora las tablas con las que se han generado dichos gráficos.

	Imágenes	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
Proyección y Retroproyección	Nº correspondencias	1168	1165	1181	32	82	168	175
	Error Correspondencias	1,531918	2,345826	1,9101333	2,2001995	35,10207	3,8172912	25,121703
	Error F	4,0484256	3,0222657	9,390209	2402,554	82,746592	479,68259	66,777473
Búsqueda Exhaustiva	Nº correspondencias	1250	960	782	24	81	169	124
	Error Correspondencias	1,4486859	2,8635284	6,4582169	49,225655	36,583234	6,1083804	67,729338
	Error F	7,0133584	49,69282	1391,8753	21723,002	1038,8745	483,93866	316,81468
Combinación de las anteriores	Nº correspondencias	1176	1102	986	21	77	162	121
	Error Correspondencias	1,3597916	2,0242833	2,3822056	66,526598	35,060742	3,2150178	70,13788
	Error F	9,5282119	3,5707866	115,44546	6495,49	651,813	479,68259	342,39548

Tabla C.9: Resultados completos del refinamiento de la matriz fundamental extraída mediante correspondencias iniciales densas con el descriptor DAISY.

	Imágenes	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
Proyección y Retroproyección	Nº correspondencias	1225	1176	1109	36	73	166	152
	Error Correspondencias	1,7795312	2,4658438	1,3964352	41,414772	192,07328	8,6046093	69,847659
	Error F	8,3055811	6,8640272	19,126703	7979,4223	732,65182	479,68259	341,13698
Búsqueda Exhaustiva	Nº correspondencias	1041	823	741	32	47	88	74
	Error Correspondencias	1,7539681	2,5542583	3,1164045	92,589429	243,58866	17,003068	67,869208
	Error F	23,094937	7,1119188	148,49266	9962,6678	1078,1074	628,2056	203,32413
Combinación de las anteriores	Nº correspondencias	1003	782	748	24	36	77	58
	Error Correspondencias	1,9310215	2,3617908	3,2070759	89,761372	265,28269	13,7317	104,35518
	Error F	19,951978	2,4519506	41,904492	6800,7398	1064,3685	563,9707	156,50459

Tabla C.10: Resultados completos del refinamiento de la matriz fundamental extraída mediante correspondencias iniciales densas con el descriptor DSIFT.

C.2.2. Errores de las correspondencias

Para comparar los descriptores utilizados en la fase de refinamiento, se han generado los gráficos a partir de los errores de las correspondencias que producen cada una de las técnicas de mejora implementadas. Esas gráficas están hechas con las tablas que se muestran a continuación.

Imágenes	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
Estimación Inicial	3,936854	41,8341358	128,224278	1118,0277	147,108129	121,427389	114,109587
Proyección y Retroproyección	3,73055397	41,7961386	127,887267	1117,49128	147,433722	121,334634	114,13307
Búsqueda Exhaustiva	2,34964	31,4342647	117,691387	1212,90722	139,159618	124,662432	110,146603
Combinación	1,77876555	50,8280992	138,658421	830,103392	149,860621	126,897021	91,0028371

Tabla C.11: Correspondencias iniciales locales y fase de refinamiento DSIFT. Gráfico 1 Figura 5.4.

Imágenes	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
Estimación Inicial	3,936854	41,8341358	128,224278	1118,0277	147,108129	121,427389	114,109587
Proyección y Retroproyección	3,90744295	41,7739596	127,823876	1118,17007	147,15108	120,954629	114,217994
Búsqueda Exhaustiva	3,19987911	31,8391096	168,662216	766,501651	198,460375	245,192004	103,594598
Combinación	3,42588769	43,4421547	60,5934941	793,181456	148,313588	159,060665	111,183706

Tabla C.12: Correspondencias iniciales locales y fase de refinamiento DAISY. Gráfico 2 Figura 5.4.

Imágenes	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
Estimación Inicial	8,37425771	6,81863592	19,5933218	8021,63382	732,413084	479,682588	345,672897
Proyección y Retroproyección	3,73055397	41,7961386	127,887267	1117,49128	147,433722	121,334634	114,13307
Búsqueda Exhaustiva	8,30558106	6,86402723	19,1267032	7979,42233	732,651825	479,682588	341,136976
Combinación	19,9519775	2,45195057	41,9044921	6800,73979	1064,36852	563,970701	156,504591

Tabla C.13: Correspondencias iniciales densas y fase de refinamiento DSIFT. Gráfico 1 Figura 5.5.

Imágenes	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
Estimación Inicial	8,37425771	6,81863592	19,5933218	8021,63382	732,413084	479,682588	345,672897
Proyección y Retroproyección	4,04842565	3,02226568	9,39020901	2402,55396	82,7465924	479,682588	66,7774728
Búsqueda Exhaustiva	7,01335844	49,6928201	1391,87533	21723,0021	1038,87447	483,938659	316,814679
Combinación	9,52821189	3,57078656	115,44546	6495,49002	651,812995	479,682588	342,39548

Tabla C.14: Correspondencias iniciales densas y fase de refinamiento DAISY. Gráfico 2 Figura 5.5.

Imágenes	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
FO SIFT	9,87429474	35,213601	15,9226119	252,120107	271,057372	220,285943	250,28603
DSIFT	1,21666399	49,6082708	19,255343	222,242477	197,295668	190,266018	233,217075
DAISY	1,56237556	36,4723642	14,6697911	249,901948	174,783493	118,410507	206,160552

Imágenes	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
FO NRDC	1,70116516	2,40412109	1,95115802	3,04783759	36,4051426	2,28803541	30,5721052
DSIFT	1,77953121	2,46584376	1,39643518	41,4147718	192,073281	8,60460933	69,8476593
DAISY	1,53191797	2,34582597	1,91013333	2,20019953	35,1020702	3,81729115	25,1217028

Tabla C.15: Errores en las correspondencias utilizando la técnica de Proyección y Retroproyección. Gráficos de la Figura 5.9

Imágenes	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
FO SIFT	9,87429474	35,213601	15,9226119	252,120107	271,057372	220,285943	250,28603
DSIFT	1,38929192	11,0567184	15,6830895	223,744332	213,483361	200,130157	205,406912
DAISY	3,11835832	15,9195606	17,7246177	204,368215	255,205377	157,123219	211,021405

Imágenes	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
FO NRDC	1,70116516	2,40412109	1,95115802	3,04783759	36,4051426	2,28803541	30,5721052
DSIFT	1,75396807	2,5542583	3,1164045	92,5894293	243,588658	17,0030683	67,8692084
DAISY	1,44868586	2,86352843	6,45821691	49,2256554	36,5832345	6,10838039	67,7293383

Tabla C.16: Errores en las correspondencias utilizando la técnica de Búsqueda Exhaustiva. Gráficos de la Figura 5.10

Imágenes	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
FO SIFT	9,87429474	35,213601	15,9226119	252,120107	271,057372	220,285943	250,28603
DSIFT	1,22766913	21,5633756	11,5591118	202,753005	144,350589	90,8134544	198,240568
DAISY	1,12060967	5,02960678	7,74320336	133,333455	201,79092	107,082377	203,467945

Imágenes	0000-0001	0000-0002	0000-0003	0000-0004	0000-0005	0000-0006	0000-0007
FO NRDC	1,70116516	2,40412109	1,95115802	3,04783759	36,4051426	2,28803541	30,5721052
DSIFT	1,93102145	2,36179075	3,20707594	89,7613725	265,282687	13,7317005	104,355175
DAISY	1,35979164	2,02428331	2,38220563	66,5265983	35,0607417	3,21501777	70,1378798

Tabla C.17: Errores en las correspondencias utilizando la técnica de Combinación de ambas. Gráficos de la Figura 5.11