

# Método automático para el incremento de correspondencias entre imágenes con perspectivas no paralelas y ampliamente separadas para su utilización en reconstrucción 3D con modelos de aprendizaje profundo

Ramiro Verrastro<sup>1</sup>, Diego Patricio Durante<sup>1</sup>, Juan Carlos Gómez<sup>1,2</sup> y Claudio A. Verrastro<sup>1,3</sup>

*1 Universidad Tecnológica Nacional, Facultad Regional Buenos Aires, Grupo de Inteligencia Artificial y Robótica (GIAR), Av. Medrano 951, (C1179AAQ) Ciudad Autónoma de Buenos Aires, Argentina*

*2 Instituto Nacional de Tecnología Industrial (INTI), Centro de Electrónica e Informática, Av. Gral. Paz 5445, (B1650WAB) San Martín, Buenos Aires, Argentina*

*3 Comisión Nacional de Energía Atómica (CNEA), Centro Atómico Ezeiza, Investigador consulto, Avenida Pbro. Juan González y Aragón (Ex Camino Real) n<sup>o</sup> 15, (B1802AYA) Ezeiza, Buenos Aires, Argentina*

*ramiroverraastro@frba.utn.edu.ar*

*Recibido el 30 de julio de 2022, aprobado el 15 de agosto de 2022*

## Resumen

En procesos que requieren realimentación visual en forma autónoma, tales como la reconstrucción tri-dimensional o el corregistro de imágenes, se necesita encontrar la mayor cantidad posible de puntos homólogos. Para ello es necesario encontrar puntos característicos en las distintas imágenes y luego hacerlos corresponder (*matching*). En el caso de que la asociación sea ideal, los puntos correspondientes son homólogos. El proceso de asociación de puntos característicos se puede dividir en 3 etapas: (1) Detección, (2) Descripción y (3) Correspondencia. En este trabajo se presenta un método que hace uso de la geometría epipolar para aumentar la cantidad de correspondencias válidas entre imágenes homólogas.

**PALABRAS CLAVE:** VISIÓN POR COMPUTADORA - CORRESPONDENCIA DE PUNTOS CARACTERÍSTICOS - ASIFT - FOTOGRAMETRÍA - GEOMETRÍA EPIPOLAR

## Abstract

In processes that require autonomous visual feedback, such as three-dimensional reconstruction or image co-registration, to find as many homologous points as possible is needed. This can be done finding keypoints in the different images and then matching them. If the association is ideal, the matched keypoints are homologous. The keypoint association process can be divided into 3 stages: (1) Detection, (2) Description, and (3) Matching. In this work, a method to increase the number of valid correspondences between homologous images using the epipolar geometry model is presented.

**KEYWORDS:** COMPUTER VISION - KEYPOINT MATCHING - ASIFT - PHOTOGRAMMETRY - EPIPOLAR GEOMETRY

## Introducción

El uso de imágenes para la medición de distancias y formas de objetos comenzó incluso antes de la aparición de la fotografía. Su utilización en topografía data de 1726 con el relevamiento del monte Pilatus en Suiza llevado a cabo, mediante perspectivas dibujadas a mano, por M. A. Carpellier. Luego, entre 1849 y 1851 Aimé Laussedat, considerado el padre de la fotogrametría, aplica los mismos conceptos de Carpellier pero ya sí, con fotografías. Él desarrolla los métodos básicos para la confección de planos (Cheli, 2011). Según Bonneval, "la fotogrametría es la técnica cuyo fin es estudiar y definir con precisión la forma, dimensiones y posición en el espacio de un objeto cualquiera, utilizando esencialmente medidas hechas sobre una o varias fotografías de ese objeto" (IGN, 2022)(Quirós Rosado, 2014).

Desde aquellos logros pioneros se desarrollaron una gran variedad de dispositivos mecánicos, ópticos, fotográficos, eléctricos y electrónicos con fines fotogramétricos. Además, con el advenimiento de las imágenes digitales y su tratamiento por *software*, el mejoramiento y la automatización de estos procesos fotogramétricos son muy significativos (Szeliski, 2022) y con, básicamente, los mismos principios. Desde muy temprano se identificó la necesidad de reconocer puntos homólogos en los cálculos de fotogrametría (Schenk, 2002) para encontrar la profundidad a partir de imágenes bidimensionales. Al principio este reconocimiento se realizaba "manualmente" y, más modernamente, en forma automática, por medio de extracción de puntos característicos (*keypoints*) usando algoritmos específicos (Tomasi 1994). Esta última técnica involucra la ubicación de un punto característico definido por algunas características (*features*) de su entorno inmediato. Estos puntos pueden ser marcas naturales o artificiales para el caso de entornos instrumentados. Sus características deben ser robustas a los cambios de perspectivas, rotación y de condiciones de iluminación. Ahora bien, este proceso de reconocer puntos homólogos de manera automática tiene tres etapas, la primera consiste en detectar puntos con características locales singulares en cada una de las vistas empleadas. La segunda describe estos puntos mediante características invariantes al cambio de perspectiva, por último, la tercera encuentra la correspondencia (*matching*) de estos puntos con sus potenciales homólogos en las otras vistas (Szeliski, 2022).

El problema de encontrar la correspondencia de puntos homólogos de una misma escena es transversal a muchas áreas de visión artificial. Esta técnica se emplea en una innumerable variedad de aplicaciones, como robótica móvil con mapeo y localización simultánea (SLAM) (Paz et al, 2008); reconstrucción 3D (Aharchi, 2020); medición de trayectoria de objetos (blancos) móviles (Szeliski, 2022); pegado de imágenes (*image stitching*) y registración (Nidhal, 2021); hasta, por supuesto, los mismos problemas iniciales de relevamiento topográfico, muchos hechos con drones (Gupta, 2018).

En este trabajo se presenta una plataforma (*framework*) para la identificación automática de puntos característicos, su descripción y el establecimiento de la correspondencia con sus potenciales homólogos en otras vistas (*matching*). En los sistemas que usan técnicas de aprendizaje profundo se requiere un número elevado de imágenes que tengan una gran cantidad de puntos homólogos identificados. Para resolver esta problemática se propone un proceso de dos etapas: el primero es a través de técnicas convencionales como *Scale Invariant Feature Transform* (SIFT) (Lowe, 2004), *Affine-SIFT* (ASIFT) (Karami, 2015) (Morel y Yu, 2009), la segunda etapa reinyecta puntos descartados por el método anterior, pero reevaluados mediante consideraciones geométricas.

En la sección siguiente se presentan una serie de técnicas que fueron utilizadas para la generación del conjunto de datos, la calibración de las imágenes, la asociación de puntos homólogos y la validación semiautomática de las correspondencias. Además, se explica en forma detallada las etapas del proceso de asociación (detección, descripción y correspondencia) y cómo se implementan los distintos filtros utilizados en la etapa de

correspondencia. En la sección de resultados se presenta la evaluación de desempeño del filtrado epipolar empleando un conjunto de datos obtenido de la bibliografía. Los resultados del incremento de la cantidad de correspondencias se muestran en función de los hiperparámetros. Finalmente, se concluye que el método presentado permite disponer de un conjunto de datos con gran cantidad de correspondencias anotadas.

## Métodos

### Generación del conjunto de datos

Con la finalidad de evaluar un método que seleccione la mayor cantidad de puntos homólogos y el menor número de falsas correspondencias, es vital contar con un conjunto de datos (*dataset*) que facilite la validación haciendo uso de propiedades de proyección de perspectivas. Para ello se creó un conjunto de datos que en este trabajo llamamos Van Gogh.

El conjunto de datos Van Gogh está constituido por 24 imágenes a colores (RGB) con una resolución de 4000 *píxeles* de ancho por 3000 *píxeles* de alto. Cada una de las imágenes ha sido generada a partir de la captura fotográfica de un objeto creado para este conjunto de datos, en el que la diferencia entre cada captura con respecto al resto es el ángulo de rotación y el vector de traslación. El mencionado objeto consiste en dos ortoedros fijados mecánicamente el uno con el otro, con distintas impresiones en cada una de sus caras planas. Todas, excepto una de las impresiones, consisten en pinturas de Vincent Van Gogh. La impresión restante se corresponde con la imagen “*reference image*” del subconjunto “Graffiti” del conjunto de datos “Oxford Dataset” (*Affine Covariant Features*, 2022)

En la Figura 1, a la izquierda se pueden ver dos imágenes arbitrarias de este conjunto de datos. Adicionalmente a estas capturas, se agregan en la Figura 1 a la derecha, tableros de calibración para disminuir los efectos de la distorsión en el proceso de captura.

### Calibración de imágenes

Si bien cualquier fotografía ideal consiste en la proyección lineal del espacio de 3 dimensiones hacia un espacio plano de 2 dimensiones, debido a los sistemas de captura aparecen distintos tipos de distorsión, haciendo que la proyección no sea lineal y presente errores (Li, 2019). Nótese que al aplicar una proyección lineal desde un espacio tridimensional a uno de 2 dimensiones (como sucedería al tomar una fotografía en condiciones ideales), una línea recta en el primer espacio debería verse como una línea recta en el segundo espacio, pero en condiciones reales, esto normalmente no sucede. La calibración de imágenes se encarga de transformar una determinada imagen capturada con distorsiones para que se parezca a la imagen sin distorsión, corrigiendo las alinealidades del sistema mediante el uso de algoritmos de calibración. A los parámetros que permiten reducir la distorsión de un sistema de captura se los denomina parámetros intrínsecos. Zhang (2000) presenta una técnica para la calibración de las



**Fig. 1.** A la izquierda dos ejemplos de imágenes del set de datos Van Gogh, a la derecha dos tableros de calibración

mismas ajustando la reproyección de las imágenes capturadas con una determinada cámara y su respectiva distorsión hacia imágenes capturadas sin distorsión. El proceso de calibración de parámetros intrínsecos se realiza a través de patrones con líneas rectas, impresos sobre caras planas que conforman imágenes similares a tableros de ajedrez, denominado “patrón de tableros de ajedrez”. La técnica en cuestión requiere de múltiples capturas de un patrón de tablero de ajedrez para, en una primera etapa, utilizar técnicas de regresión para ajustar y obtener los parámetros intrínsecos que ajustan una función de interpolación que permita reproyectar las imágenes de forma que disminuyan el error de las ubicaciones esperadas de los cuadrados del patrón, agregando el tamaño esperado de cada cuadrado. Una vez obtenidos estos parámetros, los mismos se pueden utilizar para reproyectar nuevas imágenes de cualquier objeto capturadas con la misma cámara.

## Asociación de puntos homólogos

El proceso de asociación de puntos homólogos se puede separar en 3 etapas:

- Detección de puntos característicos.
- Descripción de puntos característicos.
- Correspondencia de puntos característicos.

En la Figura 2 se puede ver un esquema simplificado de las partes del proceso.

Nótese que la naturaleza de los algoritmos y la calidad de cada una de las etapas determinan la cantidad final de puntos homólogos tanto correctamente como incorrectamente asociados. Esta cantidad está relacionada principalmente por dos factores: verdaderos positivos y falsos positivos:

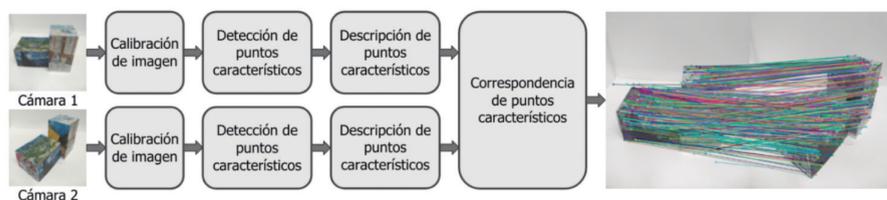
- Verdaderos Positivos *VP*: Son los puntos homólogos entre diferentes imágenes que son correctamente asociados por el sistema.
- Falsos Positivos *FP*: Son puntos asociados por el sistema, pero no son homólogos.

Para el entrenamiento de algoritmos de visión artificial resulta de interés obtener muchos puntos *VP* y muy pocos puntos *FP* para la cuantificación de la calidad del sistema se define parámetro de precisión según la ecuación 1.

$$Precision = VP / (VP + FP) \quad (1)$$

Un número elevado de puntos asociados correctamente permite que el sistema de visión artificial genere una reconstrucción con mayor granularidad, en contrapartida los *FP* deterioran la calidad de la imagen 3D resultante, generando artefactos.

Es por esto que, en este trabajo, se hace énfasis en proporcionar una herramienta que permita al diseñador resolver adecuadamente la relación de compromiso entre la cantidad de *VP* y el número de artefactos introducidos.



**Fig. 2. Diagrama de etapas simplificado de la asociación de puntos característicos**

## Detección de puntos característicos

La primera etapa de la asociación consiste en detectar puntos de manera robusta en distintas imágenes, para la correcta correspondencia en etapas posteriores, buscando la mayor cantidad de *VP*. Alineados con este fin, existen algoritmos conocidos como “Detectores de puntos característicos”. Uno de los más conocidos y pionero en la detección de puntos característicos es SIFT.

Este algoritmo pretende asociar puntos entre dos imágenes, pero con pequeños cambios de perspectiva, mediante descriptores que son invariantes a cambios de iluminación, al escalado y a la rotación (Jian Wu *et al*, 2013).

Como en este trabajo se utilizan imágenes homólogas con cambios de perspectiva importantes entre ellas, se utilizará el algoritmo ASIFT, que simula grandes cambios de perspectiva a través de transformaciones afines, para luego aplicar el algoritmo SIFT en cada una de estas transformaciones. Como resultado, estos detectores brindan por cada punto detectado, un poliedro y un ángulo de detección en un entorno de la imagen, llamada ROI o región de interés.

## Descripción de puntos característicos

De forma similar a la etapa anterior, existe un conjunto de algoritmos que describen a un punto y su ROI, mediante “Descriptores de puntos característicos”. De esta mane-

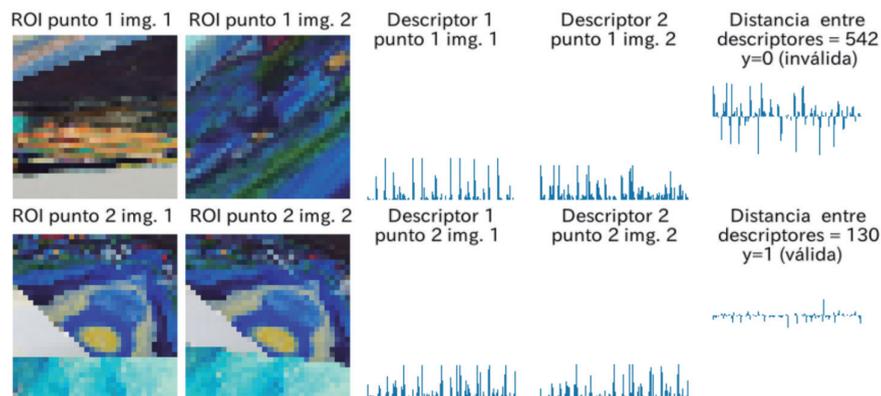


Fig. 3. Distancia entre descriptores

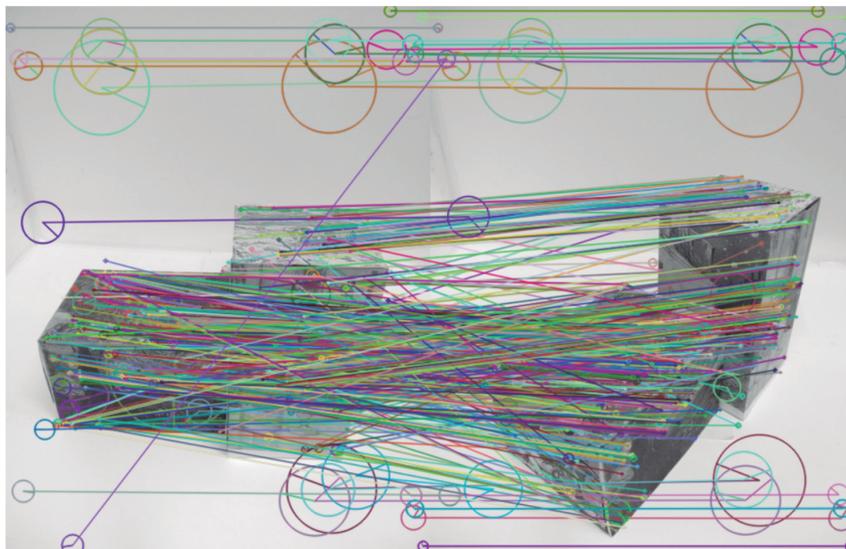


Fig. 4. Correspondencias múltiples con falsos positivos marcados con círculos

ra, cada punto puede ser representado por un vector de  $N$  elementos, que depende del algoritmo descriptor. En el presente trabajo, SIFT es utilizado como descriptor. La salida del vector brindado por SIFT consiste en un arreglo de 128 números que describen a la ROI.

## Correspondencia entre puntos característicos

El grado de correspondencia entre dos puntos de imágenes homólogas se puede estimar midiendo la distancia entre descriptores, tomados de a pares.

Una forma sencilla de comparar las distancias entre los descriptores es calculando su distancia euclídea. En el ejemplo de la Figura 3 se computa la distancia euclídea entre los descriptores ( $d_{Desc}$ ) de un par de puntos candidatos. Estableciendo un umbral se puede seleccionar el grado de similitud para el cual la asociación se considera válida o no, produciendo como salida binaria  $y=1$  (válida) o  $y=0$  (inválida).

En la Figura 4 se muestran las correspondencias (asociaciones consideradas válidas según el filtro de distancia entre descriptores) y también asociaciones inválidas (marcadas con círculos) debido a múltiples causas, como las que se enumeran a continuación:

- Deformación en las imágenes.
- Falta de robustez en los métodos de detección, descripción o evaluación de la correspondencia.
- Diferencias por la misma perspectiva.
- Oclusiones.
- Complejidad del problema por similitudes múltiples entre las imágenes.

Por estos motivos, pueden producirse *FP*, como se muestra en el ejemplo de la Figura 4.

## Validación semiautomática de las correspondencias

Dadas dos imágenes homólogas y un plano visible en ambas imágenes, a partir de la ubicación de 4 pares de puntos homólogos y conocidos, y contenidos en dicho plano, se puede estimar una matriz de transformación de perspectiva. Dicha matriz permite proyectar cualquier punto contenido dentro de esos 4 puntos desde una imagen hacia la otra, obteniendo la ubicación del mismo punto en la segunda imagen. En la Figura 5, con flechas rojas se representa la proyección de los vértices de un triángulo desde la primera imagen hacia la segunda, con una flecha verde se representa la proyección de un punto cualquiera, en el supuesto de que esté contenido en el mismo plano. De esa manera, la matriz permite estimar la ubicación de un polígono en una imagen a partir de su ubicación en la otra y usando su matriz de proyección.

Si no hay grandes cambios de perspectiva, la transformación afín resulta una buena apro-

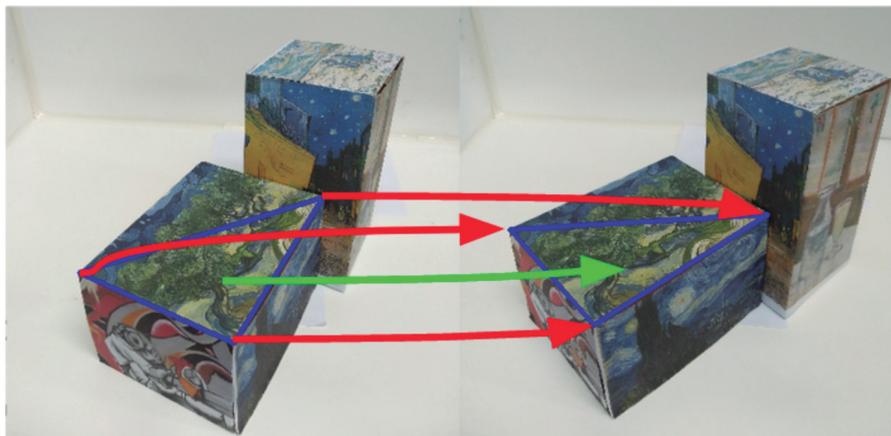
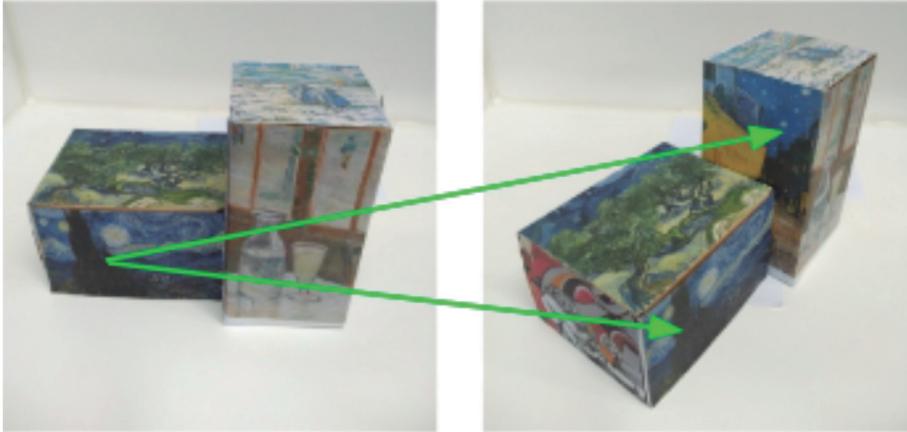


Fig. 5. Proyección de puntos entre imágenes homólogas



**Fig. 6. Filtrado por cociente de distancias entre descriptores**

ximación para la transformación de perspectiva, y la matriz de transformación se puede obtener con sólo 3 puntos homólogos. En este trabajo se utiliza dicha propiedad para calcular automáticamente la calidad de las correspondencias en las distintas etapas del sistema de la siguiente forma:

Dados un par de puntos candidatos que se sospecha que son homólogos, se obtiene la ubicación del punto de la primera imagen en la segunda. Una vez obtenida esta proyección se puede calcular, en la segunda imagen, la distancia euclídea entre el punto proyectado en la segunda imagen y la ubicación del punto candidato. Si esta distancia es menor a un umbral, podemos decir que los puntos son correspondientes. Para determinar los planos que contienen a los puntos, se utilizaron los vértices visibles de las caras de las figuras del set de datos, haciendo uso del etiquetado manual. Nótese que con algunos pocos vértices etiquetados se obtienen miles de puntos transformados y de posible validación automática.

Para mitigar los errores, disminuyendo el número de falsos positivos se agregan etapas de filtrado que se describen en las secciones siguientes.

### Filtrado por cociente de distancias entre descriptores

Debido a que algunos puntos característicos tienen ciertas similitudes entre ellos, lo mismo sucede con sus descriptores, haciendo que algunos puntos de una imagen, tenga múltiples correspondencias en la imagen homóloga, lo que puede derivar en el descarte de una asociación buena, o cometer un error por asociación falsa.

Una correspondencia está compuesta por dos puntos característicos y sus descriptores. Dada un punto en una imagen, y una lista de candidatos con posibles correspondencias en su homóloga, ordenadas por distancia, el filtrado por cociente consiste en descartar las asociaciones en donde la distancia entre el primer y segundo candidato son similares. Para optimizar esta tarea se utiliza el algoritmo de  $K$  vecinos más próximos (KNN) con  $K=2$ , con su implementación para estos fines (OpenCV FLANN, 2022).

En el ejemplo de la Figura 6 se puede observar el caso en que un punto característico de la imagen izquierda, tiene su primer y segundo candidatos en la homóloga, con una similitud tal en su entorno que hace que la distancia entre sus descriptores sea pequeña, uno de ellos produciría una correspondencia válida, en cambio el otro produciría un falso positivo. En ese caso, y sin más evidencia es necesario descartar ambos.

### Geometría epipolar

Dadas dos cámaras que captan un mismo objeto desde diferentes perspectivas, existen relaciones geométricas entre sus puntos homólogos. A esta geometría se la denomina

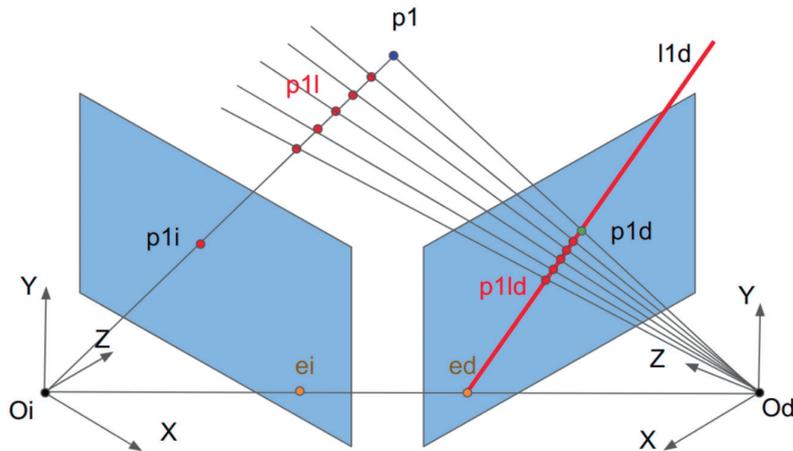
geometría epipolar (Hartley y Zisserman, 2004).

En el gráfico de la Figura 7 se presenta la relación epipolar entre dos imágenes de este tipo. Nótese que aplicando dicha geometría se restringen los posibles lugares en donde se ubica el mismo punto en las diferentes imágenes. Suponiendo que cada imagen se corresponde con la captura de una cámara, se presentan las siguientes relaciones:

- $p1$  es algún punto en el campo de visión de ambas cámaras.
- $O_i$  y  $O_d$  son los orígenes de coordenadas de las cámaras izquierda y derecha.
- $e_i$  y  $e_d$  son puntos especiales, denominados epipolos, éstos son el origen de las líneas epipolares y son fijos para una relación de perspectivas dada entre dos imágenes. En este caso su ubicación está dentro de la imagen, pero los mismos podrían estar ubicados fuera de ella.
- $p1_i$  y  $p1_d$  son las proyecciones del punto  $p1$  en las respectivas imágenes.
- $l1_d$  es la epilínea establecida a través de la geometría epipolar dado un punto proyectado  $p1_i$  en la imagen izquierda. Nótese que los distintos puntos marcados como  $p1_i$  se proyectan en distinta ubicación sobre la imagen derecha  $p1_d$  pero siempre sobre la epilínea  $l1_d$ , no obstante se proyectan en el mismo punto en la imagen izquierda  $p1_i$ .

En la Figura 8 empleando dos imágenes del conjunto de datos Van Gogh, se muestra la ubicación de un punto en la imagen izquierda y su epilínea en la imagen derecha.

Si se desea relacionar las imágenes teniendo presente la captura de las cámaras, se debe



**Fig. 7. La geometría epipolar representa la relación geométrica entre escenas similares con distinta perspectiva**



**Fig. 8. Un punto amarillo en la imagen izquierda, y su epilínea en la imagen derecha. En la geometría epipolar, se relaciona analíticamente la posición relativa entre dos puntos homólogos, a través de distintas matrices.**

considerar, además de la posición y orientación, los parámetros intrínsecos de las mismas. La matriz fundamental contiene la información sobre la transformación relativa entre dos cámaras (Luong y Faugeras, 1996).

Dicha matriz se calcula analíticamente haciendo uso de la ubicación de los puntos homólogos entre las imágenes (Hartley, 1992). En el caso de que se dispongan candidatos que potencialmente puedan ser correspondencias válidas y también *outliers*, la matriz fundamental se estima utilizando RANSAC (Yang y Li, 2013).

La ecuación 2 brinda la relación o el grado de correspondencia entre puntos de imágenes homólogas. En ella,  $x$  es la posición de un punto en la imagen derecha,  $x'$  es la posición de un punto en la imagen de la izquierda,  $F$  es la matriz fundamental, y  $D$  una magnitud que será más pequeña mientras la distancia desde los puntos hasta sus epilíneas sea más pequeña, siendo cero si los puntos pasan por dicha epilínea.

$$x'^T * F * x = D \quad (2)$$

Cabe destacar que, en el caso de cámaras estereoscópicas con perspectivas paralelas, se suele realizar una simplificación del problema rectificando las imágenes. Haciendo que cada epilínea coincida con la fila en la cual se encuentra el punto característico de la imagen homóloga, estableciendo una correspondencia unívoca entre filas de imágenes homólogas.

En este trabajo donde las perspectivas están separadas por un ángulo amplio (Pritchett y Zisserman, 1998) el proceso de rectificación no resulta adecuado, ya que se degrada notablemente la calidad de las imágenes en la proyección. Es por esto que se utilizan las imágenes sin rectificar a un único punto de vista y directamente se establecen las relaciones entre los puntos de las mismas a través de ecuaciones.

### Filtrado por geometría epipolar

Aplicar la matriz fundamental permite eliminar asociaciones de aquellos puntos, cuya distancia a la epilínea es mayor que cierto umbral, como se muestra en el gráfico de la Figura 9. Esta estrategia es robusta frente al ruido y resulta útil en problemas reales.

- $p_1$  es algún punto en la imagen de la izquierda. En este trabajo se toma el centro de la ventana en torno al punto característico, previamente detectado y descripto.
- $l_{1d}$  es la epilínea en la imagen de la derecha. Si la matriz fundamental es ideal, el punto  $p_1'$  en la segunda imagen debería estar en esta línea.
- Si existen dos candidatos para un mismo punto  $p_1$ , y sus puntos en la imagen de la derecha son  $p_1'$  y  $p_1''$ , la coincidencia debería estar ubicada idealmente sobre la epilínea  $l_{1d}$ , o en un entorno cercano.

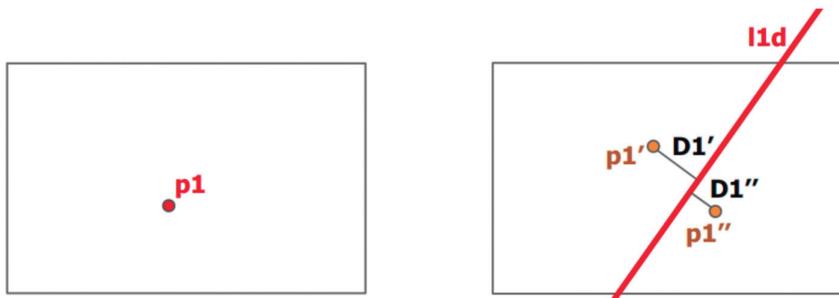


Fig. 9.  $D_1'$  y  $D_1''$  son las distancias a la epilínea  $l_{1d}$  del punto  $p_1$

- Las distancias entre ambos candidatos  $p1'$  y  $p1''$  a la epilínea  $l1d$  son  $D1'$  y  $D1''$  en la imagen de la derecha.

En la Figura 10 se muestran ejemplos de puntos y sus correspondientes epilíneas para 9 puntos, similares a los mencionados anteriormente, utilizando dos imágenes del conjunto de datos Van Gogh. Debido a que en estos casos las correspondencias son válidas, las distancias entre cada punto en la imagen izquierda y su epilínea en la imagen derecha es muy pequeña.

### Reinyección de puntos característicos

En esta sección se presenta un método de reinyección de puntos característicos, utilizando puntos descartados en etapas anteriores, en el que a través del modelo geométrico descrito previamente, es posible aumentar el número de correspondencias válidas conocidas.

Como motivación, si se toman los puntos característicos como una nube y se realiza una triangulación entre correspondencias válidas conocidas, resulta factible calcular la posición en 3 dimensiones para cada par.

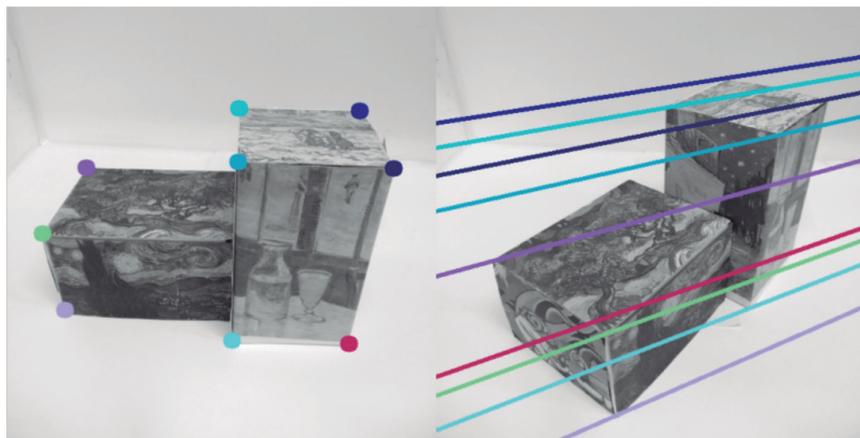
De esta forma se modela la distancia entre un punto de vista y los distintos *píxeles* o partes de la imagen. Mientras más correspondencias válidas se obtengan, será posible realizar reconstrucciones en 3 dimensiones con mayores detalles. Es por ello que se desarrolla la etapa de reinyección de puntos característicos, con la expectativa de obtener una mayor cantidad de puntos para realizar una reconstrucción más granularizada y con mayor información.

A través de la geometría epipolar, se calcula para cada caso la distancia entre un punto candidato y la epilínea correspondiente. El error máximo tolerado  $EM$  es la distancia máxima tolerada de un candidato a su epilínea correspondiente, siendo un hiperparámetro del algoritmo de reinyección.

Teniendo en cuenta el valor  $EM$  y calculando las distancias de los puntos candidatos a las epilíneas, se dan los siguientes casos:

Caso A: Solo uno de los candidatos cumple la condición de distancia, según ecuación (3), se descarta el que no la cumple.

$$(D1' < EM) \wedge (D1'' > EM) \vee (D1' > EM) \wedge (D1'' < EM) \quad (3)$$



**Fig. 10.** Puntos seleccionados en la imagen izquierda y sus líneas epipolares en la imagen derecha

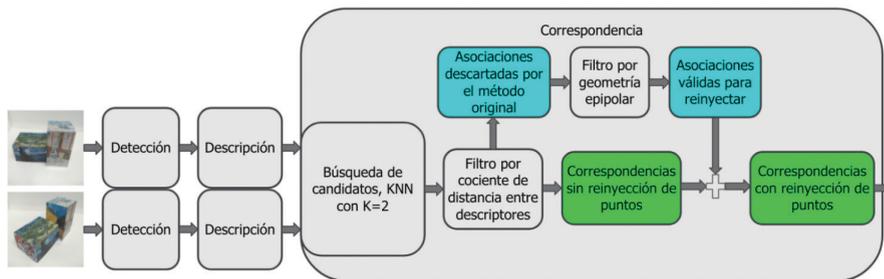


Fig. 11. Diagrama en bloques del método de reinyección de puntos característicos

Caso B: Los candidatos tienen una distancia mayor a  $EM$  y ambos se descartan.

$$(D1' > EM) \wedge (D1'' > EM) \quad (4)$$

Caso C: Según la ecuación (5), los candidatos cumplen la condición de  $EM$ . Se elige el que posea menor distancia euclídea en el espacio de los descriptores  $dDesc$ . Aplicando nuevamente el filtrado por cociente de distancia entre descriptores.

$$(D1'' < EM) \wedge (D1'' < EM) \quad (5)$$

En el diagrama de la Figura 11 se muestra como están dispuestos los filtros según el método de reinyección de puntos característicos propuesto:

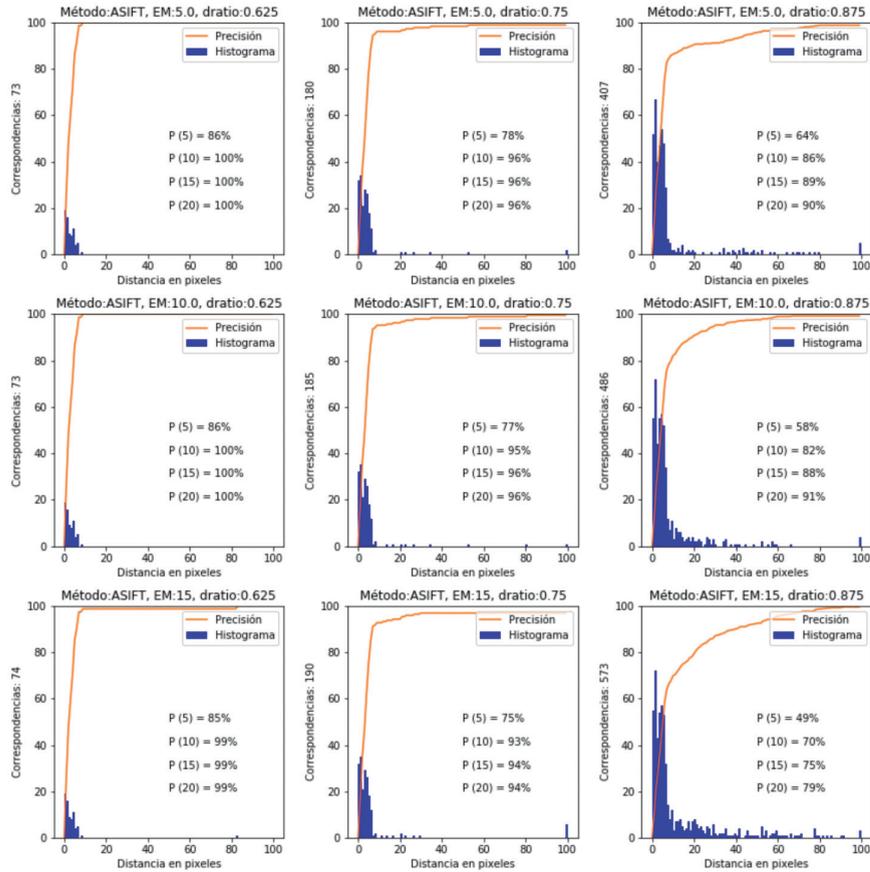
Utilizando el filtro por geometría epipolar en combinación con el filtro por cociente de distancia entre descriptores y reinyectando algunos puntos descartados, se evita descartar casos como el visto en la Figura 6, en donde si bien existen candidatos con descriptores similares dentro de la imagen, solo uno de ellos está en las cercanías de la línea epipolar correspondiente al punto bajo análisis.

Al aplicar la reinyección basada en geometría epipolar, se obtiene un incremento sustancial en la cantidad de nuevas correspondencias.

## Resultados

Con el fin de evaluar de manera estandarizada el desempeño del filtrado por geometría epipolar, se utilizan los algoritmos en imágenes pertenecientes a un conjunto de datos establecido en la bibliografía (Morel y Yu, 2009). Los algoritmos tienen hiperparámetros que deben ser ajustados de acuerdo a la aplicación. En este caso se emplea el método de selección por grilla, presentando una matriz de gráficos ordenada considerando los hiperparámetros.

En cada uno de los gráficos de la Figura 12 se muestran los resultados para un par de imágenes del conjunto de datos, utilizando el método ASIFT sin la etapa de reinyección, pero filtrando su salida por distancia a la epilínea  $EM$ . Cada gráfico presenta un histograma de las correspondencias en función de la distancia en *píxeles* del punto correspondiente hallado, respecto del punto homólogo. Para determinar si una correspondencia es válida o inválida, se establecen distintos umbrales de distancia en *píxeles*  $X$  y se calcula la precisión  $P(X)$  que también se muestra en el gráfico. La Figura 12 presenta una grilla de 9 gráficos organizados en 3 filas por 3 columnas. Cada fila tiene el mismo valor distancia máxima hasta la epilínea  $EM$  y cada columna tiene el mismo valor de cociente de distancia entre descriptores  $dRatio$ . Esta organización permite evaluar rápidamente el comportamiento de cada algoritmo en función de sus hiperparámetros de acuerdo la precisión y la granularidad requeridas por la aplicación.



**Fig. 12. Resultados para ASIFT. Donde  $P(X)$  es la precisión tomando como umbral  $X$  píxeles para considerar la asociación como válida**

De igual forma, en la Figura 13 se muestran los resultados para el algoritmo reinyectado (ASIFTR), para las mismas imágenes e iguales valores de los hiperparámetros.

En ambas grillas, se observa cómo, a medida que se relajan los requerimientos  $EM$  y  $dRatio$ , se obtiene una mayor cantidad de correspondencias a costa de una degradación en la precisión.

Los hiperparámetros mostrados en las Figuras 12 y 13 son los que brindan mejores resultados para el método ASIFT, sin embargo, como se observa, estos hiperparámetros no son los mejores para el método ASIFTR ya que producen una degradación en la precisión.

Estos hiperparámetros se deben optimizar para el caso específico de algoritmo ASIFTR. En la Figura 14, a la izquierda, se muestra un caso con los mejores hiperparámetros para ASIFT, en el centro los resultados para el método ASIFTR con los mismos hiperparámetros, y en el gráfico de la derecha los resultados con los hiperparámetros mejorados para ASIFTR.

En la Tabla 1 se comparan numéricamente los resultados, se puede observar que el método ASIFTR casi duplica el número de puntos correspondiente con un leve incremento en la precisión en todos los casos.

Esto mismo se observa en el gráfico de la Figura 15, donde se muestra que el algoritmo ASIFTR propuesto produce un mayor número de asociaciones válidas en todos los casos de umbrales de distancia en píxeles.

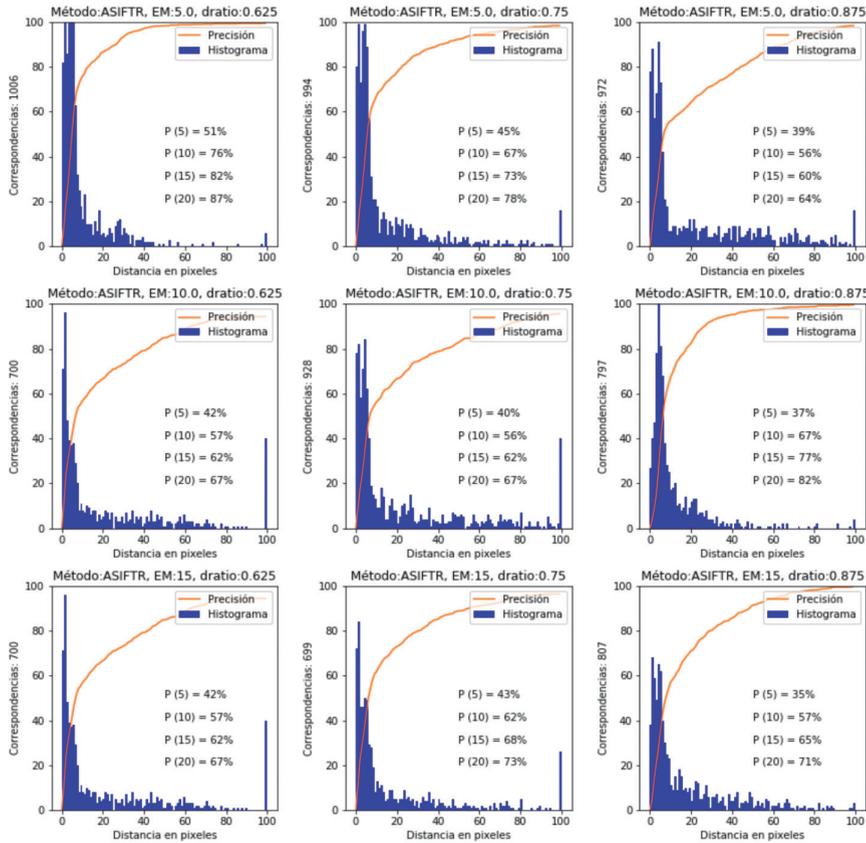


Fig. 13. Resultados para filtrado por geometría epipolar (ASIFT con reinyección de puntos). Donde P(X) es la precisión tomando como umbral X píxeles para considerar la asociación como válida

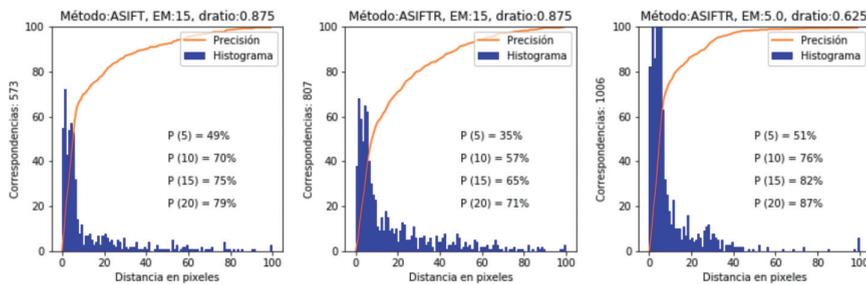


Fig. 14. Comparación de resultados entre ambos métodos. A la izquierda los mejores hiperparámetros para el Método ASIFT, en el centro ASIFTR con los mismos hiperparámetros y a la derecha el método ASIFTR con hiperparámetros mejorados

Tabla 1. Comparación de métricas de resultados entre ambos métodos

Parámetro	ASIFT (con mejores hiper parámetros para ASIFT)	ASIFTR (con mejores hiper parámetros para ASIFT)	ASIFTR (con mejores hiper parámetros para ASIFTR)
EM	15	15	5
dRatio	0,875	0,875	0,625
Correspondencias detectadas (VP + FP)	573	807	1006
P(5)	49%	35%	51%
VP con P(5)	281	279	517
FP con P(5)	292	528	489
P(20)	79%	71%	87%
VP con P(20)	454	573	871
FP con P(20)	119	234	135

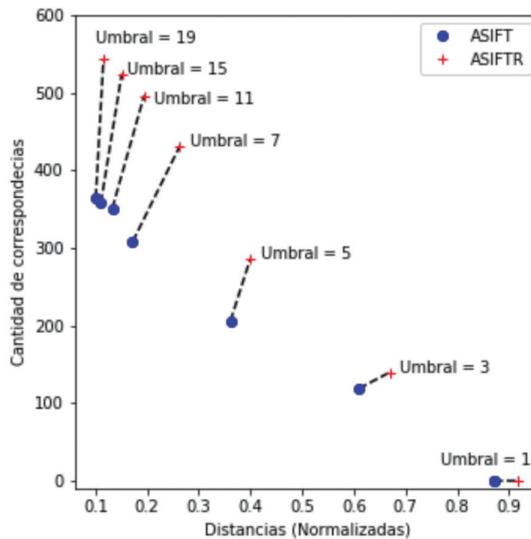


Fig. 15. Excursión permitida por los algoritmos, con distintos umbrales de distancia

## Conclusiones

El método presentado toma ventaja del modelo geométrico para aumentar la cantidad de puntos correspondientes en dos imágenes homólogas manteniendo la precisión de estas correspondencias.

Para el entrenamiento de algoritmos de visión artificial y reconstrucción en 3D con modelos de redes neuronales profundas es necesario disponer de conjunto de datos con gran cantidad de imágenes etiquetadas. Este método permite generar esas imágenes en forma automática con mayor cantidad de puntos anotados. Esto puede mejorar la granularidad de las imágenes 3D reconstruidas sin pérdida de la precisión manteniendo acotada la introducción de artefactos de reconstrucción.

## Referencias

- AFFINE COVARIANT FEATURES, (2022). Visitado 27/07/2022 <https://www.robots.ox.ac.uk/~vgg/research/affine/>
- CHELI, A.E., (2011). Introducción a la fotogrametría y su evolución – 1ª edición – La Plata, Hespérides, ISBN 978-987-1844-04-3.
- DOYLE, F., (1963). The Historical Development of Analytical Photogrammetry. The Autometric Corporation, Washington, D. C.
- GUPTA, S.K. y SHUKLA, D.P., (2018). Application of drone for landslide mapping, dimension estimation and its 3D reconstruction. *J Indian Soc Remote Sens* 46, 903–914 (2018). <https://doi.org/10.1007/s12524-017-0727-1>
- HARTLEY, R.I., (1992). Estimation of relative camera positions for uncalibrated cameras. In Proc. 2nd European Conference on Computer Vision, G. Sandini (Ed.), Santa Margherita Ligure, Italy, Springer-Verlag, vol. LNCS 588, pp. 579-587.
- HARTLEY, R. y ZISSERMAN, A., (2004). Multiple View Geometry in Computer Vision. Segunda Edición. Cambridge University Press. ISBN: 0521540518
- Instituto Geográfico Nacional, (2022). Visitado 27/07/2022 <https://www.ign.es/web/resources/docs/IGNCnig/OBS-Fotogrametria.pdf>
- WU, J. et al, (2013). A comparative study of sift and its variants. *Measurement Science Review*. DOI: 10.2478/msr-2013-0021.
- KARAMI, E.; SIVA, P. y SHEHATA, M., (2015). Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images. In: Nov. 2015.
- LI, X.; ZHANG, B.; SANDER, P.V. y LIAO, J., (2019). Blind geometric distortion correction on images through deep learning. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4855-4864
- LOWE, D., (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60 (Nov. 2004), pp. 91– DOI: 10.1023/B:VI.0000029664.99615.94.
- LUONG, Q.-T. y FAUGERAS, O. D., (1996). The Fundamental Matrix: Theory, Algorithms, and Stability Analysis. *International Journal of Computer Vision*. Vol 17. pp. 43-75. DOI: 10.1007/BF00127818.
- MONCEF, A. y M'HAMED, A.K., (2020). A Review on 3D Reconstruction Techniques from 2D Images. 10.1007/978-3-030-37629-1\_37.
- MOREL, J.M. y YU, G., (2009). ASIFT: A New Framework for Fully Affine Invariant Image Comparison, *SIAM Journal on Imaging Sciences*, vol. 2, issue 2.
- NIDHAL K. EL ABBADI et al, (2021). A Review over Panoramic Image Stitching Techniques *J. Phys.: Conf. Ser.* 1999 012115.
- OpenCV FLANN, (2022). Feature Matching with FLANN. Obtenido de [https://docs.opencv.org/3.4/d5/d6f/tutorial\\_feature\\_flann\\_matcher.html](https://docs.opencv.org/3.4/d5/d6f/tutorial_feature_flann_matcher.html)
- PAZ, L.; PINIES P.; TARDOS, J. y NEIRA, J., (2008). Large-scale 6-DOF SLAM with stereo-in hand, *IEEE Transactions on Robotics* 24 (5) 946–957.
- PRITCHETT, P. y ZISSERMAN, A., (1998). Matching and Reconstruction from Widely Separated Views. In: Koch, R., Van Gool, L. (eds) *3D Structure from Multiple Images of Large-Scale Environments*. SMILE 1998. Lecture Notes in Computer Science, vol. 1506, pp. 78-92. Springer, Berlin, Heidelberg. ISBN: 978-3-540-49437-9
- QUIRÓS ROSADO, E., (2014). Introducción a la Fotogrametría y Cartografía aplicadas a la Ingeniería Civil. Universidad de Extremadura. Servicio de Publicaciones, C/ Caldereros, 2 - Planta 2ª. 10071 Cáceres (España) ISBN de méritos: 978-84-697-1317-4. Pág 46.