

Actividad de Profundización: Algoritmos Modernos de Visión por Computador

Angel Humberto Olivera Pinzon

Objetivo General

Comprender el funcionamiento de los algoritmos de visión por computador más relevantes en la actualidad —en particular aquellos utilizados para emparejamiento de características, reconstrucción 3D y modelado neuronal de escenas— mediante análisis individual y discusión colaborativa.

Parte 1: Estudio Individual del Material de Clase

Algoritmo #1: Stitching

El stitching de imágenes se refiere al proceso de alinear, transformar y fusionar fotogramas de imágenes para producir una imagen panorámica o un mosaico de mayor tamaño.

El algoritmo busca y utiliza puntos de interés (o features) comunes en las áreas donde las imágenes se superponen. Luego, calcula la transformación geométrica necesaria (como rotación y traslación) para que esos puntos clave coincidan perfectamente, y finalmente combina las imágenes para crear un resultado sin costuras visibles.

Tipos Comunes de Transformación

El corazón del algoritmo de stitching reside en el cálculo de la matriz de transformación (H) que mapea los píxeles de una imagen a la otra. Las más comunes son:

1. Transformación Proyectiva (Homografía): Es la transformación más común y robusta. Permite proyectar imágenes tomadas desde diferentes perspectivas o ángulos (por ejemplo, una cámara giratoria). Se representa como una matriz 3×3 que mapea un plano a otro.

2. Transformación Afín: Una transformación más simple, adecuada para imágenes donde la perspectiva no cambia drásticamente. Mantiene el paralelismo de las líneas.
3. Rígida: Solo permite rotación y traslación (movimiento sin deformación).

¿Qué problema resuelve?

El principal problema que resuelve el algoritmo de stitching es la limitación del campo de visión (FOV) de una única cámara.

Los problemas específicos que el stitching aborda son:

1. Campo de Visión Limitado

- Problema: Una sola toma con una cámara estándar solo puede capturar una porción limitada de una escena grande (como un paisaje, un mural o el interior de un edificio).
- Solución: Al tomar múltiples fotos solapadas y unirlas, el stitching permite generar una imagen panorámica que representa la escena completa, superando la limitación física del hardware.

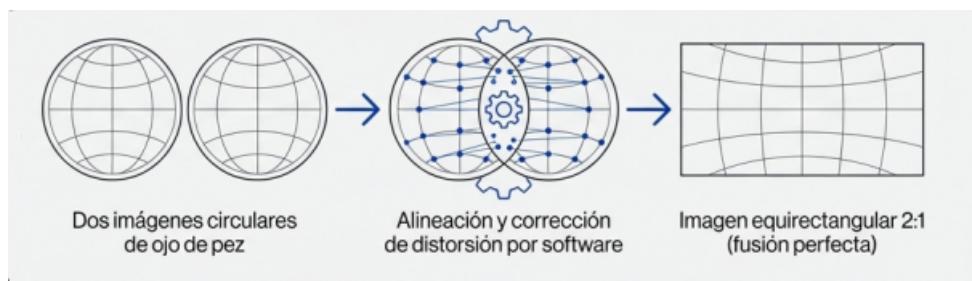
2. Resolución Insuficiente

- Problema: A veces, necesitas capturar una escena grande con un nivel de detalle muy alto (por ejemplo, para cartografía o inspección industrial). Si te alejas para capturar todo en una sola toma, pierdes resolución.
- Solución: El stitching de muchas imágenes de primer plano (cada una de alta resolución) crea un mosaico final con una resolución total mucho mayor que la imagen que podría producir un solo sensor.

3. Falta de Contexto Espacial

- Problema: En aplicaciones como la inspección aérea con drones o la robótica (SLAM), es crucial tener un mapa visual coherente del entorno. Las imágenes individuales son fragmentadas.
- Solución: Al coser las imágenes, se crea un mapa visual unificado y contextual, permitiendo la localización, la navegación y el mapeo.

¿Cuál es su flujo de operación (pipeline)?

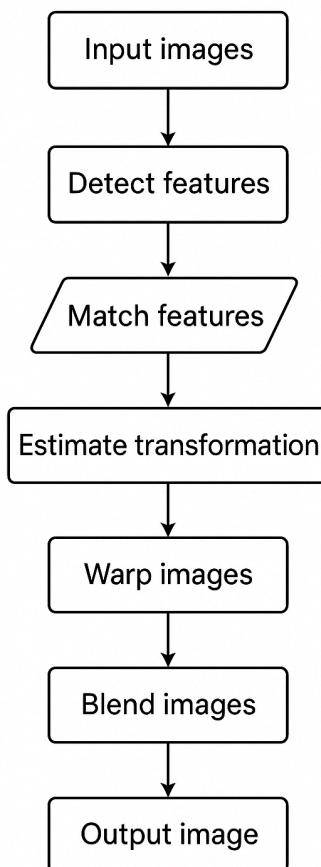


¿Qué ventajas presenta frente a técnicas tradicionales?

Ventaja	Descripción	¿Por qué es Superior a Métodos Manuales?
Precisión Geométrica	Utiliza técnicas avanzadas (como la Homografía) basadas en puntos clave (SIFT, ORB) para calcular la transformación exacta.	Es imposible lograr a mano la precisión sub-píxel necesaria para alinear perspectivas. Los métodos manuales solo permiten traslación simple, no corrección de perspectiva.
Invariancia a la Perspectiva	El uso de la Homografía permite alinear correctamente imágenes tomadas desde diferentes ángulos y distancias de una escena plana.	Los métodos manuales (o el simple blending sin transformación) fallan estrepitosamente cuando hay un cambio de perspectiva.
Fusión (Blending) Suave	Implementa algoritmos de blending avanzados (como la Laplacian Pyramid) para suavizar las diferencias de color, brillo y exposición en las áreas de solapamiento.	Si solo se "pega" o superpone una imagen encima de otra, las líneas de corte, las diferencias de iluminación y las "costuras" son claramente visibles.

Robustez al Ruido (Outliers)	Emplea métodos como RANSAC (RANdom SAmple Consensus) en la fase de emparejamiento para descartar automáticamente los matches de características incorrectos causados por reflejos, ruido o puntos ambiguos.	Las técnicas manuales o primitivas no tienen un mecanismo para identificar y corregir errores de emparejamiento.
Automatización y Velocidad	El proceso completo (detección, emparejamiento, estimación de la matriz y blending) se realiza automáticamente, típicamente en segundos.	El stitching de múltiples imágenes de forma manual es extremadamente tedioso y lento, incluso para usuarios expertos.

Diagrama de flujo



Algoritmo #2: ORB

ORB fue desarrollado por Ethan Rublee, Vincent Rabaud, Kurt Konolige y Gary R. Bradski en 2011. Combina y mejora dos algoritmos preexistentes:

1. FAST (Features from Accelerated Segment Test): Se utiliza para la detección eficiente de los puntos clave (corners) en una imagen. FAST es conocido por su alta velocidad de cómputo.
2. BRIEF (Binary Robust Independent Elementary Features): Se utiliza para la descripción de los puntos clave detectados. BRIEF es un descriptor binario que también es muy rápido de calcular y comparar.

ORB toma la velocidad de FAST y de BRIEF, pero añade dos mejoras cruciales que abordan las limitaciones de sus predecesores:

- Orientación (Oriented): Asigna una orientación a cada punto clave detectado por FAST. Esto se logra calculando el centro de masa del parche alrededor del punto.

- Rotación (Rotated): Modifica el descriptor BRIEF para que sea invariante a la rotación. Esto significa que el descriptor será el mismo incluso si la imagen está rotada.

¿Qué problema resuelve?

El problema fundamental que resuelve ORB, y los algoritmos de detección y descripción de características en general, es la identificación y comparación robusta de los mismos puntos o regiones en diferentes imágenes, a pesar de variaciones como la rotación, el cambio de escala, la iluminación o el ruido.

Específicamente, ORB busca resolver las siguientes limitaciones de sus algoritmos base:

1. Falta de Invariancia a la Rotación (Limitación de FAST y BRIEF):
 - Problema: Los detectores y descriptores originales como FAST y BRIEF no se desempeñan bien cuando una imagen está rotada, lo que es común en aplicaciones del mundo real.
 - Solución de ORB: Al añadir la orientación y rotar los patrones de muestreo del descriptor BRIEF, ORB se vuelve invariante a la rotación, permitiendo emparejar características incluso si el objeto o la cámara han rotado.
2. Lento Tiempo de Cómputo (Limitación de SIFT/SURF):
 - Problema: Algoritmos más antiguos y robustos como SIFT (Scale-Invariant Feature Transform) y SURF (Speeded Up Robust Features) son muy precisos y son invariantes a la escala y rotación, pero son computacionalmente muy costosos y lentos para aplicaciones en tiempo real (como robótica móvil o realidad aumentada).
 - Solución de ORB: ORB es significativamente más rápido de calcular y emparejar (al usar un descriptor binario como BRIEF), lo que lo hace ideal para aplicaciones que requieren alta velocidad sin sacrificar demasiada precisión. A menudo se le describe como una alternativa de alto rendimiento en tiempo real a SIFT y SURF.

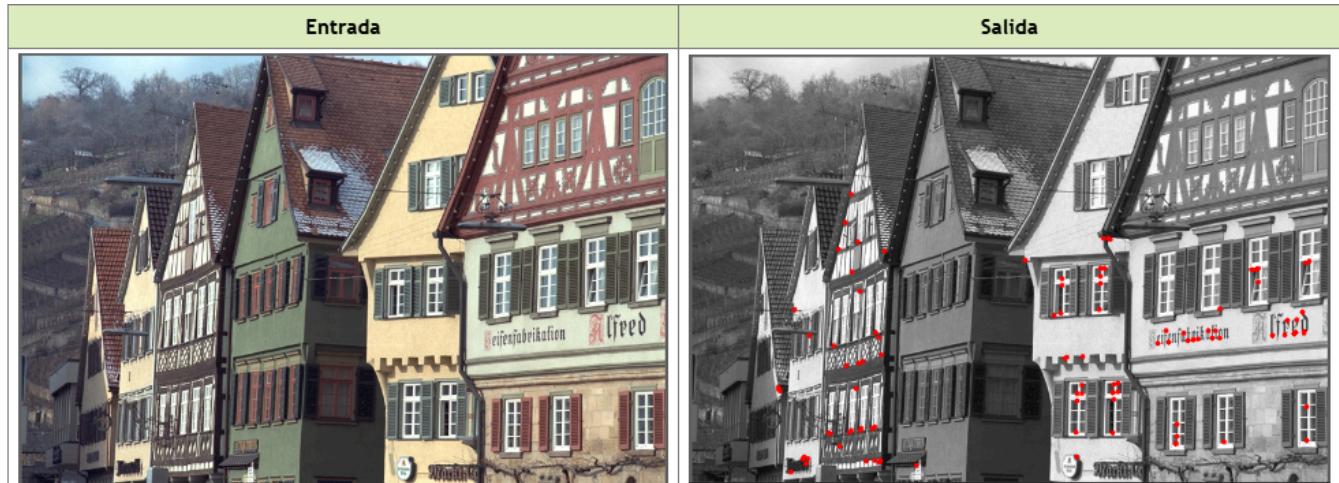
Aplicaciones Comunes

Gracias a su velocidad y robustez, ORB se utiliza en:

- Stitching de Imágenes (Image Stitching): Para alinear y fusionar múltiples imágenes en un panorama.
- Reconocimiento de Objetos (Object Recognition): Para identificar objetos en diferentes poses.

- Seguimiento y Mapeo (SLAM - Simultaneous Localization and Mapping): En robótica y drones para que el dispositivo sepa dónde está y mapee su entorno simultáneamente.
- Realidad Aumentada (AR): Para rastrear marcadores o puntos clave en el mundo real para superponer contenido digital.

¿Cuál es su flujo de operación (pipeline)?



¿Qué ventajas presenta frente a técnicas tradicionales?

Las técnicas tradicionales de detección y descripción de características incluyen los detectores como Harris y FAST (sin orientación) y los descriptores como SIFT (Scale-Invariant Feature Transform) y SURF (Speeded Up Robust Features).

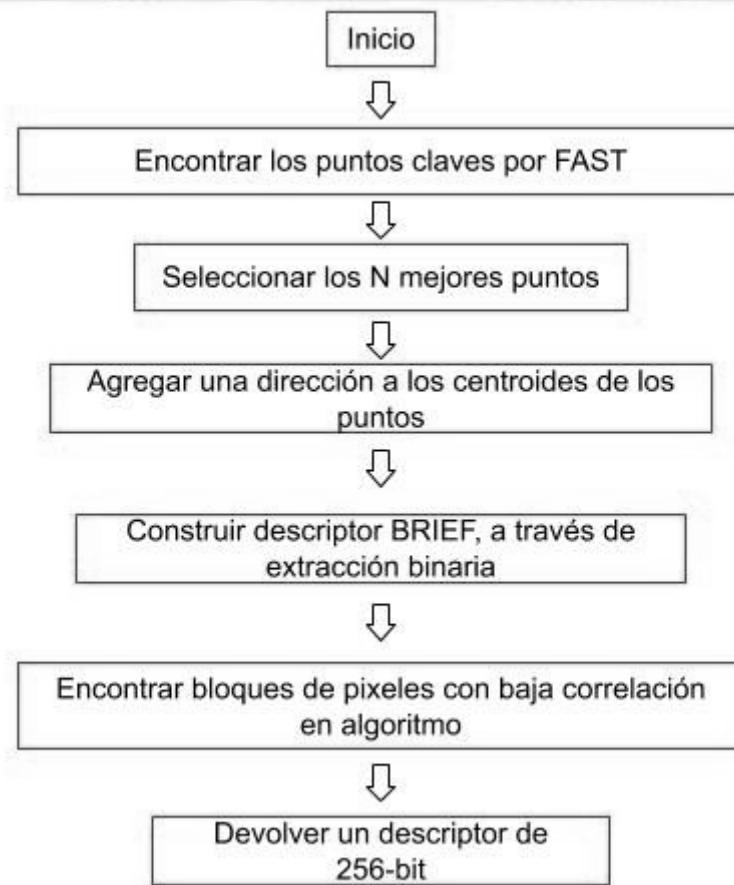
ORB está diseñado específicamente para combinar la robustez de SIFT/SURF con la velocidad de FAST/BRIEF, ofreciendo las siguientes ventajas clave:

Ventaja	Explicación	Frente a SIFT/SURF	Frente a FAST/BRIEF (no orientados)

Alta Velocidad	El uso de FAST para la detección y BRIEF (binario) para la descripción garantiza un tiempo de cómputo extremadamente bajo.	Mucho más rápido para la detección y descripción. Ideal para aplicaciones en tiempo real.	Igual de rápido en detección, pero el descriptor es más rápido de emparejar (distancia de Hamming).
Invariancia a la Rotación	Al añadir la orientación y usar rBRIEF, el descriptor es robusto a la rotación de la imagen.	Comparable en robustez a la rotación.	Superior. Los descriptores BRIEF puros fallan si la imagen está rotada.
Descriptor Compacto	El descriptor binario ocupa menos espacio en memoria y el emparejamiento es muy rápido con la distancia de Hamming.	El descriptor SIFT/SURF es un vector de números de punto flotante de alta dimensión (e.g., 128 o 64 elementos), lo que requiere más memoria y cálculos más lentos.	Superior en robustez (rBRIEF está mejor optimizado para varianza).
Open Source	ORB es de código abierto (publicado bajo la licencia BSD) y no tiene costos de licencia.	SIFT y SURF tenían patentes asociadas (aunque ya expiraron), lo que limitaba su uso en proyectos comerciales sin licencia.	—

Diagrama de flujo

Diagrama de Flujo del Algoritmo ORB



Algoritmo #3: SIFT

SIFT es un algoritmo patentado (originalmente por la Universidad de Columbia Británica y desarrollado por David Lowe en 1999) utilizado para la detección y descripción de puntos de interés locales (o características clave) en una imagen.

Lo que distingue a SIFT y lo hizo un estándar durante muchos años es que las características que detecta y describe son invariantes a la escala y la rotación. Esto significa que:

1. Invariancia a la Escala: Un objeto puede ser detectado si está cerca (grande) o lejos (pequeño).
2. Invariancia a la Rotación: Un objeto puede ser detectado incluso si está girado en cualquier ángulo.

El descriptor SIFT es un vector de características de 128 elementos que resume la distribución de gradientes de intensidad en la región alrededor del punto clave.

¿Qué problema resuelve?

El problema central que resuelve SIFT es la identificación y el emparejamiento fiable de objetos o escenas entre diferentes imágenes que presentan grandes variaciones en la perspectiva de la cámara, la iluminación, la escala y la rotación.

Antes de SIFT, los métodos de detección de esquinas (como Harris) eran sensibles a:

- Cambios de Escala: Si acercabas o alejabas la cámara, el mismo punto clave podía desaparecer o cambiar drásticamente su descripción.
- Cambios de Rotación: Si el objeto giraba, el detector y el descriptor fallaban en el emparejamiento.

SIFT resuelve este problema al garantizar que los puntos clave y sus descriptores asociados permanezcan constantes (invariantes) bajo estas transformaciones.

¿Cómo Resuelve el Problema?

SIFT introduce un flujo de trabajo de cuatro pasos (que se detallará en la siguiente pregunta, si es necesario) que aborda las limitaciones de las técnicas previas:

1. Invariancia a la Escala (Mediante Pirámide de Gaussianas): SIFT utiliza una pirámide de imágenes con diferentes desenfoques gaussianos (DOG - Difference of Gaussians) para buscar puntos clave a múltiples escalas. Esto garantiza que los puntos sean intrínsecamente estables a cualquier nivel de ampliación o reducción.
2. Invariancia a la Rotación (Mediante Asignación de Orientación): SIFT analiza las magnitudes de los gradientes alrededor del punto clave para determinar una orientación principal. Todos los cálculos de los descriptores posteriores se realizan en un marco de coordenadas rotado de acuerdo a esta orientación.
3. Robustez a la Iluminación y Perspectiva: El descriptor se construye utilizando histogramas de gradientes, que son relativamente insensibles a pequeños cambios en la iluminación o en la perspectiva tridimensional.

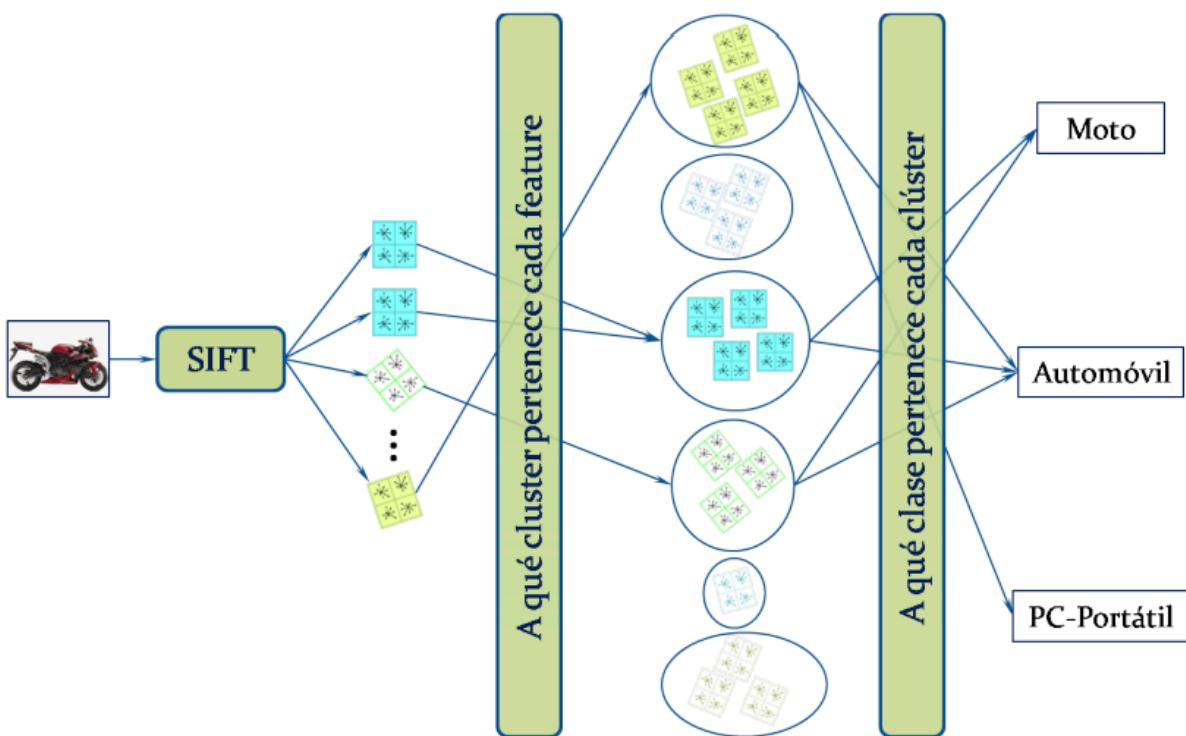
Aplicaciones Comunes

SIFT se utiliza donde la precisión y la robustez son más importantes que la velocidad:

- Stitching de Imágenes: Para alinear imágenes con precisión y crear panoramas.
- Reconocimiento de Objetos: Para identificar objetos complejos en escenas desordenadas y con variaciones de pose.

- Modelado 3D (Structure from Motion): Para reconstruir modelos 3D a partir de múltiples imágenes 2D.
- Robótica/Navegación: Donde se requiere un emparejamiento de características extremadamente fiable.

¿Cuál es su flujo de operación (pipeline)?



¿Qué ventajas presenta frente a técnicas tradicionales?

La principal ventaja de SIFT radica en la invariancia que logra para sus características.

Las técnicas tradicionales (como el detector de Harris) son detectores de esquina rápidos y precisos, pero fallan cuando la imagen cambia de tamaño o rota. El detector de Harris detectará una esquina en una imagen, pero si se acerca la cámara (cambio de escala) o se gira el objeto (cambio de rotación), el punto puede desaparecer o su descripción cambiará drásticamente.

SIFT supera esto al generar características que son robustas e invariantes a las dos transformaciones más comunes:

Invariancia	Solución de SIFT	Resultado
Escala	Busca características en múltiples escalas (pirámide de Gaussianas).	El mismo punto clave es detectado y descrito de manera similar, sin importar cuán cerca o lejos esté el objeto.
Rotación	Asigna una orientación principal al punto clave.	El descriptor se calcula en un marco de coordenadas rotado, haciéndolo inmune a la rotación del objeto o la cámara.
Iluminación	Utiliza gradientes de intensidad en lugar de valores de píxeles puros.	Es más tolerante a cambios en la iluminación y ruido.

Diagrama de flujo

