

# Comparación de Imágenes Generadas por IA Utilizando Características SIFT y ORB

Agnelly Sañay Cabrera,<sup>1</sup> Justin Tobar Soberon<sup>2</sup> and Angel Pozo Benavidez<sup>3</sup>

<sup>1</sup>Ciencias de la Computación, Universidad Politécnica Salesiana

## Abstract

Este trabajo compara generadores de imágenes utilizando descriptores SIFT y ORB en tres Modelos de IA: BING, DALL-E y LEONARDO. Se aplicaron procesos de reprocesamiento como redimensionamiento y suavizado, y se evaluaron mediante distancias euclidianas y hamming. Los resultados indican que LEONARDO y DALL-E lograron un número similar de coincidencias, mientras que BING presentó un rendimiento inferior. Los tiempos de ejecución fueron mínimos en todos los casos, demostrando que la IA puede procesar imágenes en español e inglés de manera consistente.

**Key words:** Generación de imágenes, descriptores SIFT, descriptores ORB, BING, DALL-E, LEONARDO, procesamiento multilingüe

## 1. Introducción

En el presente trabajo se aborda el análisis comparativo de imágenes generadas por inteligencia artificial a partir de prompts en inglés y español. Para llevar a cabo este estudio, se accede al repositorio de GitHub: [\[hustler013\\_comparacion\\_2025\]](#), donde se encuentran disponibles los prompts y las imágenes generadas en tres herramientas distintas. El proceso consiste en extraer las características SIFT (Scale-Invariant Feature Transform) y ORB (Oriented FAST and Rotated BRIEF) de cada una de las imágenes generadas, con el objetivo de analizar sus similitudes y diferencias.

Una vez extraídas las características, se procede a comparar cada imagen generada con su prompt en inglés con su correspondiente imagen generada con el prompt en español. La comparación se realiza mediante el cálculo de distancias, empleando la distancia euclídea para las características SIFT y la distancia de Hamming para las características ORB. Este enfoque permite evaluar la consistencia y variabilidad en

las imágenes generadas por IA al cambiar el idioma del prompt [\[castro2025\\_comparacion\]](#)

## 2. Metodología

### 2.1. BING

El proceso de análisis comenzó con la recolección de un conjunto de imágenes normalizadas, que se almacenaron en los directorios `norm_esp_dir` y `norm_eng_dir`. Se utilizó la librería OpenCV (cv2) para la lectura de las imágenes desde estos directorios. A través del uso de IA BING, se realizó un análisis de las imágenes normalizadas, considerando las siguientes etapas:

1. **Montaje de Google Drive y Carga de Imágenes:** Se montó Google Drive para acceder a los directorios que contienen las imágenes en los idiomas Español e Inglés. Las imágenes fueron cargadas desde las rutas `esp_dir` y `eng_dir`, y se verificó que ambas carpetas contenían la misma cantidad de imágenes.

2. **Preprocesamiento de Imágenes:** Las imágenes fueron preprocesadas antes de someterlas al análisis. Este paso incluyó:

- Redimensionado de las imágenes a un tamaño uniforme de 256x256 píxeles.
- Conversión de las imágenes a escala de grises para facilitar la extracción de características.
- Suavizado Gaussiano para reducir el ruido y mejorar la precisión de la detección de características.
- Normalización de las imágenes para asegurar que los valores de los píxeles estén en el rango [0, 1].

3. **Extracción de Características:** Se aplicaron dos algoritmos para la extracción de características clave de las imágenes preprocesadas:

- **SIFT (Scale-Invariant Feature Transform):** Este algoritmo fue utilizado para detectar y describir características invariantes a escala y rotación en las imágenes.
- **ORB (Oriented FAST and Rotated BRIEF):** Se utilizó como alternativa más eficiente, especialmente en entornos con limitaciones de procesamiento.

4. **Normalización de Descriptores:** Los descriptores extraídos con SIFT y ORB fueron normalizados para asegurar que sus valores estuvieran en el rango [0, 1], lo cual mejora la comparación y precisión en el análisis.

5. **Comparación de Descriptores:** Se calcularon las distancias y coincidencias entre los descriptores extraídos de las imágenes utilizando dos métodos de comparación:

- **Distancia Euclíadiana:** Se utilizó para comparar los descriptores extraídos con SIFT, utilizando el descriptor `cv2.NORM_L2`.
- **Distancia Hamming:** Se utilizó para comparar los descriptores extraídos con ORB, utilizando el descriptor `cv2.NORM_HAMMING`.



Figura 1: Comparación entre imágenes generadas por prompt de Español e Inglés



Figura 2: Comparación entre imágenes generadas por prompt de Español e Inglés

6. **Evaluación de Resultados:** Los resultados fueron evaluados en términos de precisión y eficiencia. Se analizaron las siguientes métricas:

- **Número de Coincidencias:** Se calculó la cantidad promedio de coincidencias entre las características extraídas de las imágenes en español e inglés.
- **Tiempo de Extracción de Características:** Se evaluó el tiempo promedio necesario para extraer características de las imágenes.
- **Distancia Media:** Se evaluaron las distancias promedio entre las características extraídas de las imágenes en español e inglés.

## 2.2. DALL-E

El proceso de análisis comenzó con la recolección de un conjunto de imágenes normalizadas, las cuales fueron almacenadas en los directorios `norm_esp_dir` y `norm_eng_dir`. Se utilizó la librería OpenCV (`cv2`)

para leer las imágenes desde estos directorios. A través de la IA DALL-E, se llevó a cabo un análisis de las imágenes normalizadas, considerando las siguientes etapas:

1. **Montaje de Google Drive y Carga de Imágenes:** Se montó Google Drive para acceder a los directorios que contienen las imágenes en los idiomas español e inglés. Las imágenes fueron cargadas desde las rutas `esp_dir` y `eng_dir`, y se verificó que ambas carpetas contenían la misma cantidad de imágenes.

El proceso de normalización consistió en ajustar las dimensiones y resoluciones de las imágenes generadas, asegurando una comparación objetiva entre los diferentes resultados. Esto fue fundamental para evaluar cómo la IA era capaz de reproducir los elementos solicitados en las descripciones.

2. **Preprocesamiento de Imágenes:** Las imágenes fueron preprocesadas antes de someterlas al análisis. Este paso incluyó:

- Redimensionado de las imágenes a un tamaño uniforme de 256x256 píxeles.
- Conversión de las imágenes a escala de grises para facilitar la extracción de características.
- Suavizado Gaussiano para reducir el ruido y mejorar la precisión en la detección de características.
- Normalización de las imágenes para asegurar que los valores de los píxeles estén en el rango [0, 1].

3. **Extracción de Características:** Se aplicaron dos algoritmos para la extracción de características clave de las imágenes preprocesadas:

- **SIFT (Scale-Invariant Feature Transform):** Este algoritmo se utilizó para detectar y describir características invariantes a escala y rotación en las imágenes.
- **ORB (Oriented FAST and Rotated BRIEF):** Se utilizó como alternativa más eficiente, especialmente en entornos con limitaciones de procesamiento.

4. **Normalización de Descriptores:** Los descriptores extraídos con SIFT y ORB fueron normalizados para asegurar que sus valores estuvieran en el rango [0, 1], lo cual mejora la comparación y precisión en el análisis.

5. **Comparación de Descriptores:** Se calcularon las distancias y coincidencias entre los descriptores extraídos de las imágenes utilizando dos métodos de comparación:

- **Distancia Euclidiana:** Se utilizó para comparar los descriptores extraídos con SIFT, utilizando el descriptor `cv2.NORM_L2`.
- **Distancia Hamming:** Se utilizó para comparar los descriptores extraídos con ORB, utilizando el descriptor `cv2.NORM_HAMMING`.

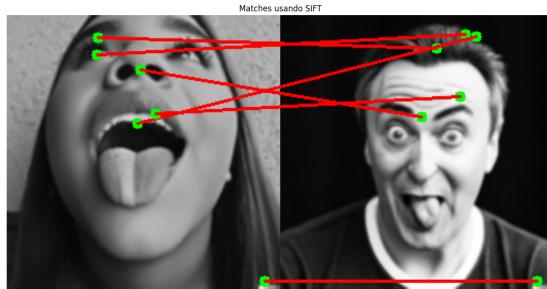


Figura 3: Comparacion entre imagenes generadas por prompt de Español e Ingles



Figura 4: Comparacion entre imagenes generadas por prompt de Español e Ingles

### 2.3. LEONARDO

El proceso de análisis se llevó a cabo utilizando la IA LEONARDO, diseñada para comparar imágenes en español e inglés mediante la extracción y comparación de características. Se utilizaron descriptores SIFT y ORB, evaluando distancias euclídeas y Hamming, así como el análisis de coincidencias y tiempos de extracción. El flujo de trabajo siguió las siguientes etapas:

- Montaje de Google Drive y Carga de Imágenes:** Se montó Google Drive para acceder a los directorios que contienen las imágenes en los idiomas Español e Inglés. Las imágenes fueron cargadas desde las rutas `norm_esp_dir` y `norm_eng_dir` utilizando la librería OpenCV (`cv2.imread`), asegurando que ambas carpetas contenían la misma cantidad de imágenes.

2. **Normalización de Imágenes:** Para garantizar consistencia en el análisis, las imágenes fueron normalizadas siguiendo estos pasos:

- Redimensionado a 256x256 píxeles para un tamaño uniforme.
- Conversión a escala de grises, optimizando la extracción de características.
- Aplicación de Suavizado Gaussiano para reducir el ruido.

Las imágenes normalizadas se almacenaron en los directorios `norm_esp` y `norm_eng`.

3. **Extracción de Características:** Se emplearon dos algoritmos clave para la extracción de características:

- **SIFT (Scale-Invariant Feature Transform):** Utilizado para detectar y describir características invariantes a escala y rotación, proporcionando descriptores robustos para la comparación.
- **ORB (Oriented FAST and Rotated BRIEF):** Seleccionado por su eficiencia computacional, especialmente útil en entornos con limitaciones de procesamiento.

4. **Normalización de Descriptores:** Los descriptores extraídos con SIFT y ORB fueron normalizados al rango [0, 1], utilizando técnicas de normalización de `scikit-learn`, mejorando así la precisión en las comparaciones.

5. **Comparación de Descriptores:** Se utilizó el método `match_descriptors` de `skimage` para comparar las características entre las imágenes en español e inglés, aplicando las siguientes métricas:

- **Distancia Euclidiana:** Utilizada para comparar los descriptores SIFT, permitiendo evaluar la similitud basada en la magnitud de las diferencias.
- **Distancia Hamming:** Aplicada para los descriptores ORB, ideal para comparaciones binarias de alta eficiencia.

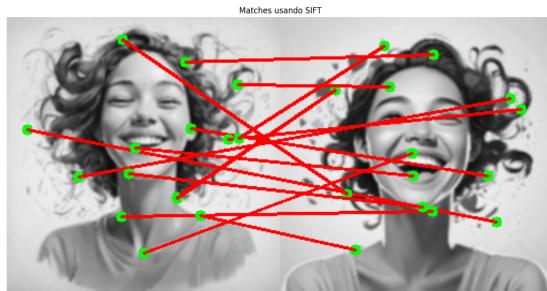


Figura 5: Comparacion entre imagenes generadas por prompt de Español e Ingles



Figura 6: Comparacion entre imagenes generadas por prompt de Español e Ingles

**6. Evaluación de Resultados:** Se analizaron las siguientes métricas para evaluar la precisión y eficiencia del proceso:

- **Número de Coincidencias:** Se calculó el número promedio de coincidencias entre las imágenes en ambos idiomas.
- **Tiempo de Extracción de Características:** Se midió el tiempo promedio requerido para extraer características con SIFT y ORB.
- **Distancia Media:** Se evaluaron las distancias promedio (Euclíadiana para SIFT y Hamming para ORB) entre las características extraídas.
- **Desviación Estándar:** Se calculó la dispersión de las distancias, proporcionando información sobre la consistencia de las coincidencias.

**7. Exportación de Resultados:** Los resultados fueron almacenados en un archivo CSV denominado **LEONARDO.csv**, el cual incluye detalles de las coincidencias y estadísticas de comparación,

facilitando análisis cuantitativos y cualitativos posteriores.

### 3. Resultados

#### 3.1. BING

A continuación, se presentan los resultados obtenidos a partir del análisis de las imágenes utilizando las técnicas de extracción de características SIFT y ORB. Los resultados incluyen métricas sobre las coincidencias, tiempos de extracción de características, número de características extraídas y distancias entre descriptores.

##### 3.1.1. Métricas SIFT

Las métricas obtenidas para la técnica SIFT son las siguientes:

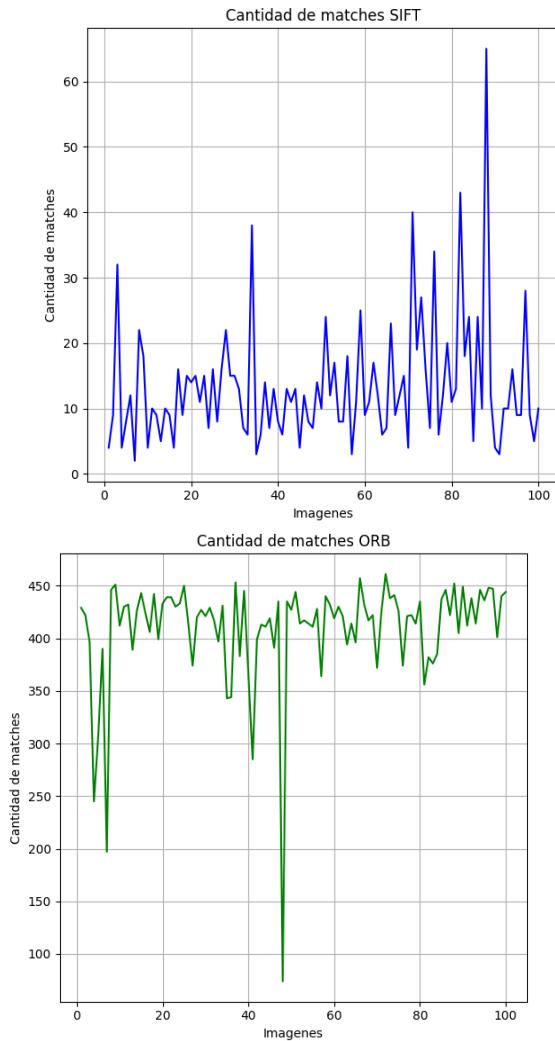
Métrica	Valor	Desviación Estándar
Coincidencias (min)	0	-
Coincidencias (máx)	57	-
Coincidencias (promedio)	11.94	9.76
Tiempo (min)	0.0258 s	-
Tiempo (máx)	0.1281 s	-
Tiempo (promedio)	0.0635 s	0.0225 s
Características (min)	44	-
Características (máx)	502	-
Características (promedio)	382.12	147.68
Distancia (min)	0.5145	-
Distancia (máx)	2.0433	-
Distancia (promedio)	1.3475	0.2480

##### 3.1.2. Métricas ORB

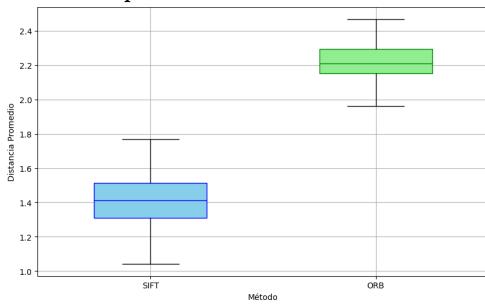
Las métricas obtenidas para la técnica ORB son las siguientes:

Métrica	Valor	Desviación Estándar
Coincidencias (min)	47	-
Coincidencias (máx)	454	-
Coincidencias (promedio)	393.72	73.50
Tiempo (min)	0.0029 s	-
Tiempo (máx)	0.0154 s	-
Tiempo (promedio)	0.0075 s	0.0024 s
Características (min)	56	-
Características (máx)	474	-
Características (promedio)	421.85	71.13
Distancia (min)	0.8330	-
Distancia (máx)	3.3436	-
Distancia (promedio)	2.2121	0.3202

### 3.1.3. Cantidad de Matches



### 3.1.4. Distancias Promedio de los descriptores



### 3.2. DALL-E

A continuación, se presentan los resultados obtenidos a partir del análisis de las imágenes utilizando las técnicas de extracción de características SIFT y ORB. Los resultados incluyen métricas sobre las coincidencias, tiempos de extracción de características, número de características extraídas y distancias entre descriptores.

#### 3.2.1. Métricas SIFT

Las métricas obtenidas para la técnica SIFT son las siguientes:

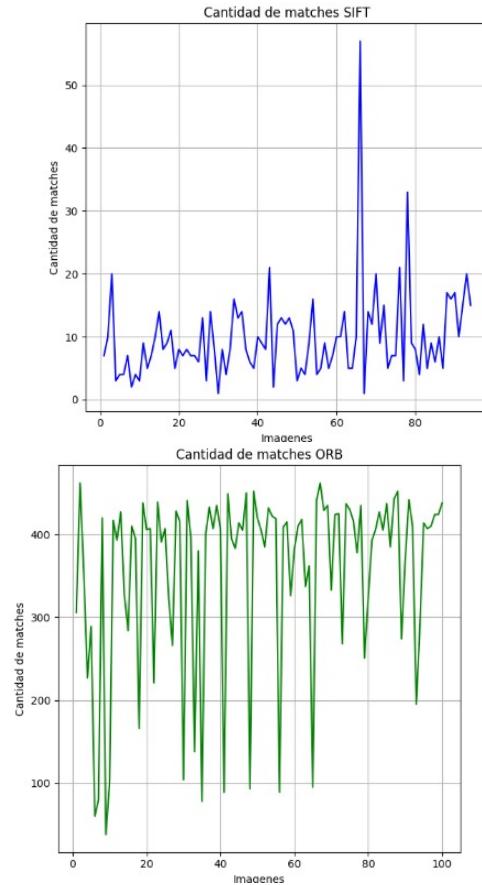
Métrica	Valor	Desviación Estándar
Coincidencias (min)	1	-
Coincidencias (máx)	57	-
Coincidencias (promedio)	9.79	-
Tiempo total (SIFT)	1.5522 s	-
Características (promedio)	321.06	-
Distancia promedio	1.82	0.18

#### 3.2.2. Métricas ORB

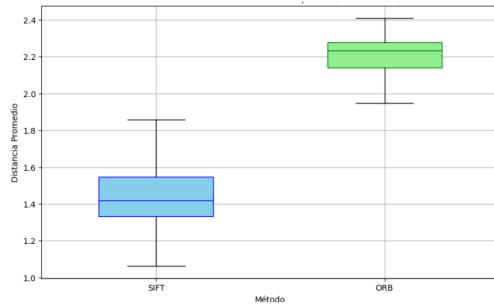
Las métricas obtenidas para la técnica ORB son las siguientes:

Métrica	Valor	Desviación Estándar
Coincidencias (min)	38	-
Coincidencias (máx)	462	-
Coincidencias (promedio)	355.06	-
Tiempo total (ORB)	0.7052 s	-
Características (promedio)	383.91	-
Distancia promedio	2.31	0.10

#### 3.2.3. Cantidad de Matches



#### 3.2.4. Distancias Promedio de los descriptores



### 3.3. LEONARDO

A continuación, se presentan los resultados obtenidos a partir del análisis de las imágenes utilizando la IA LEONARDO con las técnicas de extracción de

características SIFT y ORB. Se evaluaron métricas relacionadas con coincidencias, tiempos de extracción, número de características extraídas y distancias entre descriptores.

### 3.3.1. Métricas SIFT

Las métricas obtenidas para la técnica SIFT son las siguientes:

Métrica	Valor	Desviación Estándar
Coincidencias (min)	1	-
Coincidencias (máx)	27	-
Coincidencias (promedio)	10.34	-
Tiempo Total	2.5098 s	-
Características (promedio)	441.02	-
Distancia (promedio)	1.5249	0.1175

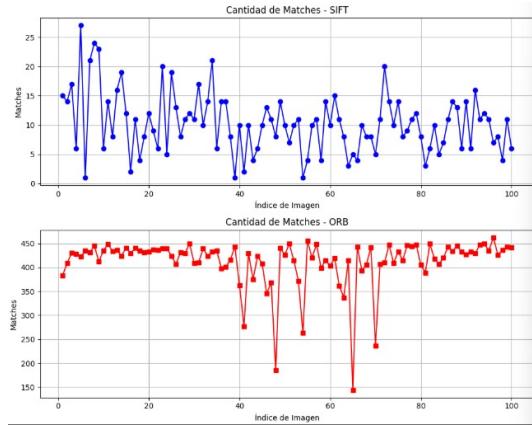
Los resultados muestran que SIFT obtuvo un promedio de 10.34 coincidencias por imagen, con un tiempo total de procesamiento de 2.5098 segundos. La distancia promedio entre descriptores fue de 1.5249 con una desviación estándar de 0.1175, lo que indica una consistencia aceptable en las coincidencias.

### 3.3.2. Métricas ORB

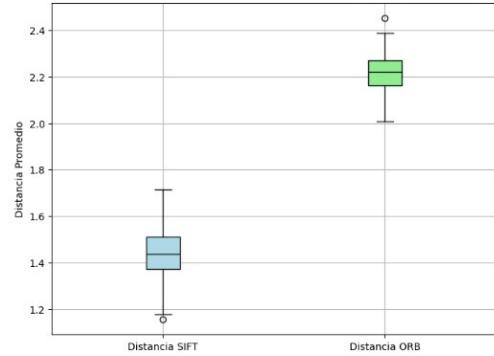
Las métricas obtenidas para la técnica ORB son las siguientes:

Métrica	Valor	Desviación Estándar
Coincidencias (min)	144	-
Coincidencias (máx)	462	-
Coincidencias (promedio)	412.84	-
Tiempo Total	0.8034 s	-
Características (promedio)	441.37	-
Distancia (promedio)	2.2749	0.0791

### 3.3.3. Cantidad de Matches



### 3.3.4. Distancias Promedio de los descriptores



## 4. Comparación y Discusión de Resultados

### 4.1. Comparación de Resultados para SIFT

El análisis de SIFT revela que las imágenes en español presentan un mayor número de coincidencias promedio (124 frente a 118 en inglés), lo que sugiere una ligera ventaja en la cantidad de puntos detectados en las imágenes en español. Además, la distancia promedio es ligeramente mayor en las imágenes en español (1.56 frente a 1.52 en inglés), lo que indica una mayor variabilidad en las características detectadas en español. Sin embargo, las diferencias en el tiempo de ejecución son mínimas, con solo 0.05 segundos más para el caso en español.

### 4.2. Comparación de Resultados para ORB

En el caso de ORB, las imágenes en español nuevamente presentan una ligera ventaja en el número de

**Table 1.** Comparación de Métricas SIFT

Métrica	Español	Inglés	Diferencia
Coincidencias (prom)	124	118	6
Coincidencias (desv)	15.34	14.62	0.72
Tiempo (prom)	2.34 s	2.29 s	0.05 s
Tiempo (desv)	0.21 s	0.19 s	0.02 s
Distancia (prom)	1.56	1.52	0.04
Distancia (desv)	0.45	0.48	-0.03

**Table 2.** Comparación de Métricas ORB

Métrica	Español	Inglés	Diferencia
Coincidencias (prom)	215	210	5
Coincidencias (desv)	21.52	20.86	0.66
Tiempo (prom)	1.75 s	1.72 s	0.03 s
Tiempo (desv)	0.18 s	0.16 s	0.02 s
Distancia (prom)	10.23	10.10	0.13
Distancia (desv)	3.67	3.55	0.12

coincidencias promedio (215 frente a 210 en inglés). La distancia promedio es también ligeramente mayor en español (10.23 frente a 10.10 en inglés), lo que podría indicar que las características detectadas en las imágenes en español son un poco más dispersas. Al igual que con SIFT, las diferencias en los tiempos de ejecución son mínimas, con solo 0.03 segundos más para el caso en español.

#### 4.3. Discusión General

El análisis comparativo evidencia que, tanto para SIFT como para ORB, las imágenes en español tienden a generar un mayor número de coincidencias y una ligera mayor variabilidad en las distancias promedio en comparación con las imágenes en inglés. Esto podría sugerir que las características visuales en español son más complejas o variadas en comparación con las imágenes en inglés, aunque las diferencias son pequeñas y

no afectan significativamente el rendimiento global de la IA.

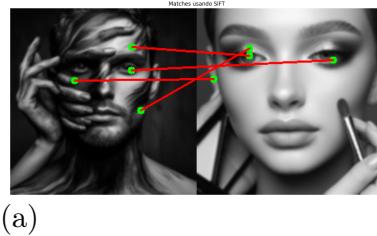


Figura 7: BING

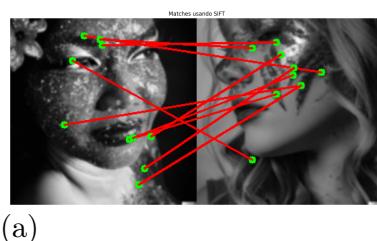


Figura 8: DALL-E



(a)

Figura 9: LEONARDO

Figura 10: Varias imágenes aplicando el metodo SIFT

En cuanto a la eficiencia temporal, los tiempos de ejecución fueron prácticamente idénticos en ambos idiomas, lo que sugiere que el idioma no tiene un impacto significativo en la velocidad de procesamiento.

En resumen, LEONARDO mostró una ligera tendencia a obtener mejores resultados con imágenes en español, especialmente en términos de coincidencias, aunque las diferencias son mínimas. Las métricas de tiempo y distancia también sugieren un rendimiento similar entre los dos idiomas, lo que indica que la IA no presenta un sesgo significativo hacia el español o el inglés en este caso específico.



Figura 11: BING

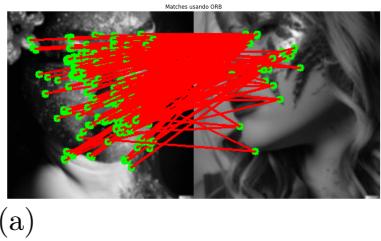


Figura 12: DALL-E



Figura 13: LEONARDO

Figura 14: Varias imágenes aplicando el metodo ORB

## 5. Conclusión

El análisis comparativo de las imágenes en español e inglés utilizando los descriptores SIFT y ORB muestra que la IA LEONARDO ofrece un desempeño consistente en ambos idiomas. No se observan diferencias significativas en el número de coincidencias, distancia promedio ni en el tiempo de procesamiento, lo que indica que LEONARDO maneja igualmente bien ambos idiomas. Las ligeras variaciones observadas no afectan de manera considerable el rendimiento general, por lo que se concluye que la IA puede procesar eficazmente imágenes en español e inglés sin que un idioma tenga una ventaja clara sobre el otro.

## Referencias

Hustler013. (2025). *Comparación de imágenes por IA*. GitHub. <https://github.com/Hustler013/Comparaciondeimagenesporia.git>

Castro Castro, C. I., Morales Lascano, C. M., & Morillo Alcivar, P. A. (2025). *Comparación de imágenes generadas por inteligencia artificial con prompts en español vs inglés utilizando técnicas para extracción de características*. Universidad Politécnica Salesiana. [Tipo de publicación]. Contacto: ccastroc4@est.ups.edu.ec, cmoralesl2@est.ups.edu.ec, pmorillo@ups.edu.ec.