

# Análisis comparativo de imágenes generadas por IA con prompts en español e inglés usando los Descriptores 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 estudio analiza las diferencias en la generación de imágenes mediante IA cuando se usan prompts en español e inglés. Se emplearon modelos de IA como Disco Diffusion, Midjourney y Stable Diffusion para generar imágenes en distintas categorías. Posteriormente, se aplicaron técnicas de extracción de características con SIFT y ORB para evaluar similitudes y diferencias estructurales entre las imágenes generadas. Los resultados indican variaciones en la cantidad y distribución de características detectadas, lo que sugiere una posible influencia del idioma en la interpretación del prompt por parte de la IA.

**Key words:** fleneración de imágenes, inteligencia artificial, prompts en español, prompts en inglés, extracción de características, SIFT, ORB, Disco Diffusion, Midjourney, Stable Diffusion, análisis comparativo.

## 1. Introducción

La inteligencia artificial ha revolucionado la forma en que creamos imágenes, permitiéndonos generar escenas increíbles a partir de simples descripciones de texto. Modelos como Disco Diffusion, Midjourney y Stable Diffusion han demostrado ser herramientas poderosas, capaces de interpretar nuestras ideas y convertirlas en imágenes detalladas y creativas. Estos avances han abierto nuevas posibilidades para artistas, diseñadores y creadores de contenido, facilitando la producción de obras visuales en cuestión de minutos. Sin embargo, sabemos que muchas de estas IA han sido entrenadas principalmente con datos en inglés, lo que nos lleva a preguntarnos si interpretarían un prompt en español de la misma manera. A medida que la globalización y la diversificación cultural continúan expandiéndose, es crucial investigar cómo estos modelos de IA responden a diferentes idiomas

y dialectos. Las diferencias culturales, idiomáticas y contextuales pueden influir en la forma en que se interpretan los mensajes, lo que a su vez podría afectar la calidad y la relevancia de las imágenes generadas. ¿Las imágenes generadas con un mismo mensaje en ambos idiomas serán idénticas o presentarán diferencias? Para responder a esta pregunta, en este estudio comparamos imágenes creadas con prompts en español e inglés, utilizando herramientas avanzadas de análisis como SIFT y ORB. Estas herramientas nos permiten detectar patrones y similitudes entre las imágenes, facilitando una comparación objetiva y precisa.[García & Torres, 2023■]

## 2. Metodología

El estudio se llevó a cabo en tres fases principales:

- Generación de imágenes.
- Extracción de características.
- Comparación de Imágenes
- Análisis comparativo.

### 2.1. Generación de imágenes.

Se generaron un total de 600 imágenes, dividiéndose en 100 imágenes en español y 100 en inglés para cada modelo de IA utilizado: Disco Diffusion, Midjourney y Stable Diffusion. La generación de imágenes se realizó utilizando un conjunto de prompts cuidadosamente seleccionados para evaluar diferentes categorías visuales.

Cada prompt fue escrito en español y traducido al inglés para asegurar la comparabilidad entre los resultados obtenidos en ambos idiomas. Las categorías seleccionadas fueron:

#### Paisaje - Landscapes:

- Español: Crea una imagen de una puesta de sol en una playa tropical con palmeras y aguas cristalinas."
- Inglés: Create an image of a sunset on a tropical beach with palm trees and crystal-clear water."

#### Arte - Art

- Español: Crea una representación artística de la 'Mona Lisa' de Leonardo Da Vinci en estilo abstracto."
- Inglés: Create an artistic representation of Leonardo Da Vinci's 'Mona Lisa' in an abstract style."

#### Rostro - Faces

- Español: Crea una imagen de un rostro humano que exprese felicidad y alegría."
- Inglés: Create an image of a human face expressing happiness and joy."

Para *Disco Diffusion* se utilizó el modelo preentrenado `sd-dreambooth-library/disco-diffusion-style`. Se implementó una función para generar imágenes, donde cada prompt fue procesado y almacenado en directorios específicos según la categoría y el idioma. Se utilizó `DiffusionPipeline` para cargar el modelo y generar las imágenes en formato PNG.

Para *Stable Diffusion* se implementó `Stable Diffusion`, se empleó el modelo `stable-diffusion-v1-5`. Se definieron funciones para organizar las carpetas de salida y generar las imágenes mediante `StableDiffusionPipeline`. Las imágenes se almacenaron en directorios

organizados por categoría e idioma.

Con *Midjourney* se realizó utilizando el modelo `prompthero/openjourney`, cargado con autenticación en Hugging Face. Se aplicó un proceso similar al de los otros modelos, donde cada imagen generada fue guardada en su respectiva carpeta según la categoría visual y el idioma del prompt.

Las imágenes generadas se almacenaron en Google Drive en una estructura:

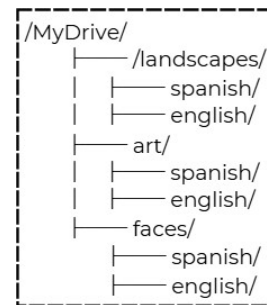


Figura 1: Estructura

### 2.2. Extracción de características

Para analizar las similitudes entre las imágenes generadas en español e inglés, se emplearon dos métodos de extracción de características: *SIFT* (Scale-Invariant Feature Transform) y *ORB* (Oriented FAST and Rotated BRIEF).

**Método SIFT - Distancia Euclidiana:** SIFT se utilizó para detectar puntos clave en las imágenes y generar descriptores robustos a cambios de escala, iluminación y rotación. Para comparar las imágenes generadas con prompts en diferentes idiomas, se calcularon las distancias euclidianas entre los descriptores de cada par de imágenes correspondientes. Proceso:

- Conversión de las imágenes a escala de grises para una mejor detección de características.
- Extracción de puntos clave y descriptores mediante el algoritmo SIFT.
- Emparejamiento de descriptores entre imágenes con el mismo identificador en español e inglés.
- Cálculo de la distancia euclidiana promedio entre los descriptores emparejados.
- Registro del número total de coincidencias encontradas.



Figura 2: Arte - Disco Diffusion



Figura 3: Rostro - Disco Diffusion

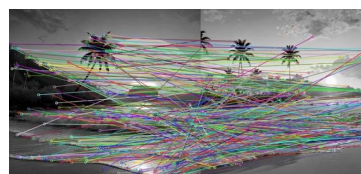


Figura 4: Paisaje - Disco Diffusion



Figura 5: Arte - Stable Diffusion



Figura 6: Rostro - Stable Diffusion

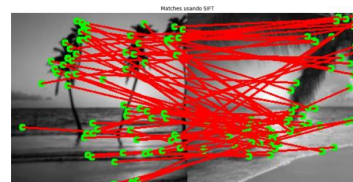


Figura 7: Paisaje - Stable Diffusion

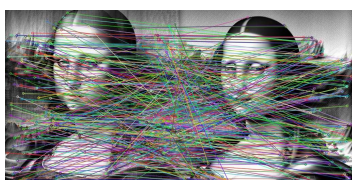


Figura 8: Arte - Midjourney



Figura 9: Rostro - Midjourney

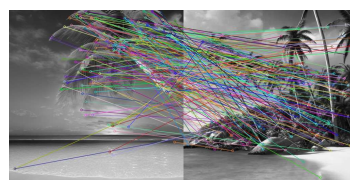


Figura 10: Paisaje - Midjourney

**Método ORB - Distancia Hamming:** ORB es un algoritmo más eficiente computacionalmente en comparación con SIFT y se basa en el uso de descriptores binarios. En este caso, la similitud entre las imágenes se evaluó mediante la distancia de Hamming entre los descriptores de cada par de imágenes correspondientes. Proceso:

- Conversión de las imágenes a escala de grises.
- Detección de puntos clave y extracción de descriptores mediante ORB.
- Emparejamiento de descriptores entre imágenes.
- Cálculo de la distancia Hamming promedio entre los descriptores emparejados.
- Registro del número total de coincidencias encontradas.

Tras procesar las imágenes con ambos métodos, se calcularon estadísticas clave para evaluar la similitud entre las imágenes generadas en español e inglés:

- Distancia máxima, mínima y media de los descriptores.
- Desviación estándar de las distancias, indicando la variabilidad en las diferencias entre imágenes.
- Cantidad máxima, mínima y promedio de coincidencias detectadas entre imágenes.
- Número promedio de características detectadas en cada imagen.



Figura 11: Arte - Disco Diffusion



Figura 12: Rostro - Disco Diffusion



Figura 13: Paisaje - Disco Diffusion



Figura 14: Arte - Stable Diffusion

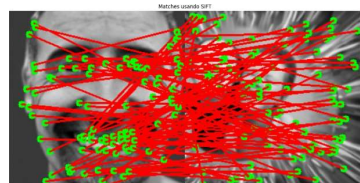


Figura 15: Rostro - Stable Diffusion

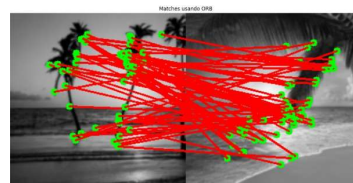


Figura 16: Paisaje - Stable Diffusion



Figura 17: Arte - Midjourney



Figura 18: Rostro - Midjourney

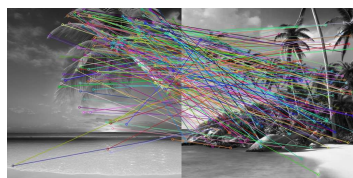


Figura 19: Paisaje - Midjourney

### 2.3. Comparación de Imágenes

Una vez extraídas las características mediante los métodos SIFT y ORB, se procedió a la comparación de las imágenes generadas en español e inglés. El objetivo de esta comparación fue evaluar el grado de similitud entre las imágenes creadas a partir de los mismos prompts en diferentes idiomas y analizar si existen diferencias significativas en la interpretación de los modelos de IA.

Para cada par de imágenes correspondientes (una generada con el prompt en español y otra con el mismo prompt en inglés):

- Emparejamiento de características: Se identificaron los puntos clave en ambas imágenes y se emparejaron los descriptores más similares.
- Cálculo de distancias: En SIFT, la similitud se midió mediante la distancia euclidiana entre los descriptores emparejados. En ORB, la similitud

se evaluó utilizando la distancia de Hamming, que mide la diferencia entre los descriptores binarios.

- Conteo de coincidencias: Se registró el número total de matches detectados entre las imágenes en español e inglés.
- Visualización de resultados: Se generaron imágenes resaltando las coincidencias detectadas entre los puntos clave de cada par de imágenes comparadas.

**Resultados:** Los resultados obtenidos permitieron analizar cómo los modelos de IA generan imágenes en función del idioma del prompt. A través de las métricas extraídas, se evaluaron los siguientes aspectos:

- Nivel de similitud entre imágenes: Cuanto menor sea la distancia promedio entre descriptores, mayor es la similitud entre las imágenes generadas en ambos idiomas.
- Diferencias en la cantidad de características detectadas: Un mayor número de matches indica que

la estructura visual de ambas imágenes es más parecida.

- Variaciones según la categoría: Se analizaron diferencias en la similitud de imágenes en función de la categoría (arte, rostros y paisajes).
- Diferencias entre modelos de IA: Se comparó el desempeño de Disco Diffusion, MidJourney y Stable Diffusion en la generación de imágenes a partir de los mismos prompts en diferentes idiomas.

2.4. Análisis Comparativo

Para analizar y comparar las imágenes generadas, se utilizó Google Colab como entorno de ejecución. Esto facilitó el procesamiento de un gran volumen de imágenes y la aplicación de algoritmos de detección de características mediante OpenCV.

El propósito del análisis fue identificar diferencias y similitudes entre las imágenes creadas por modelos de IA a partir de prompts en español e inglés. Para ello, se emplearon dos técnicas de extracción y comparación de características:

- SIFT (Scale-Invariant Feature Transform) con distancia euclidiana.
- ORB (Oriented FAST and Rotated BRIEF) con distancia de Hamming.

Ambos métodos permitieron medir la similitud visual entre pares de imágenes generadas con el mismo prompt en distintos idiomas.

1. **Comparación por modelo de IA** Se analizaron los resultados obtenidos para cada modelo de IA:

Disco Diffusion:

- **SIF**

Métrica	Valor	Desviación Estándar
Distancia máxima	308.92	-
Distancia mínima	228.40	-
Distancia media	268.62	-
Máximo matches	1257	12.37
Mínimo matches	142	-
Promedio de matches	608.66	
Tiempo total	377.90 s	
Promedio de características	2525.44	991.49

- **ORB**

Métrica	Valor	Desviación Estándar
Distancia máxima	71.19	-
Distancia mínima	52.69	-
Distancia media	62.52	-
Máximo matches	179	2.87
Mínimo matches	93	-
Promedio de matches	139.75	
Tiempo mínimo	0.1273s	
Tiempo máximo	1.5104 s	
Promedio de características	499.86	0.81

- **Diagrama**

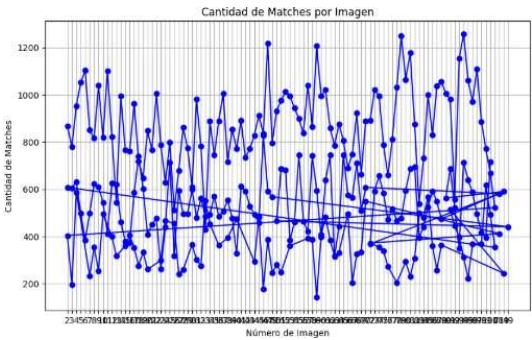


Figura 20: Cantidad de Matches - SIFT



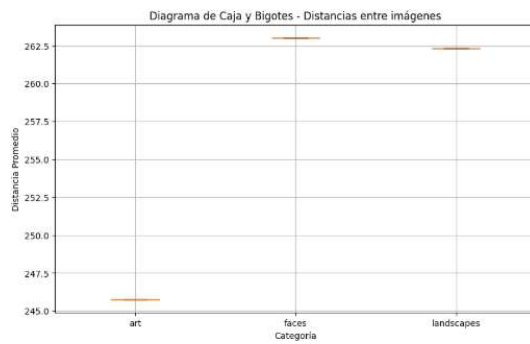


Figura 21: Distancia entre imágenes - SIFT

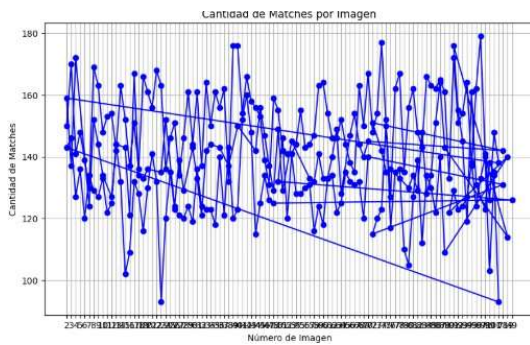


Figura 22: Cantidad de Matches - ORB

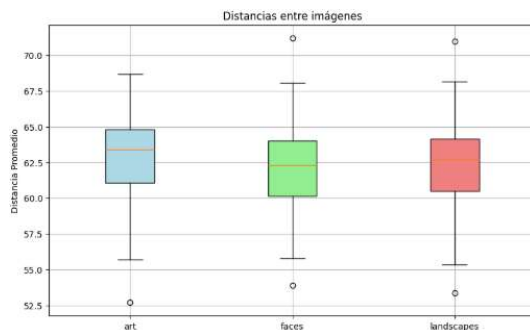


Figura 23: Distancia entre las imágenes - ORB

**Midjourney:**▪ **SIFT**

Métrica	Valor	Desviación Estándar
Distancia máxima	328.31	17.83
Distancia mínima	233.42	-
Distancia media	266.59	-
Máximo matches	1019	2.87
Mínimo matches	53	-
Promedio de matches	351.28	
Tiempo total	291.71s	
Promedio de características	1480.73	926.43

▪ **ORB**

Métrica	Valor	Desviación Estándar
Distancia máxima	73.10	-
Distancia mínima	48.00	-
Distancia media	57.58	3.48
Máximo matches	177	
Mínimo matches	92	-
Promedio de matches	132.14	
Tiempo mínimo	0.0908s	
Tiempo máximo	0.3327 s	
Promedio de características	499.86	0.81

## DIAGRAMA

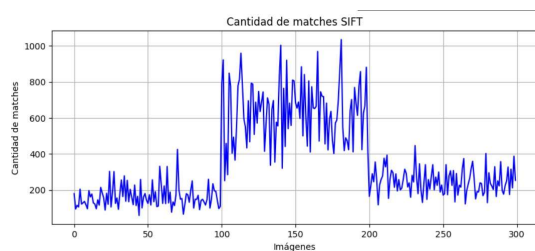


Figura 24: Cantidad de Matches - SIFT

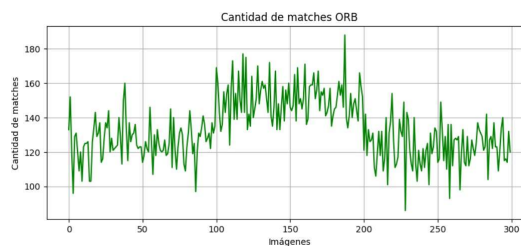


Figura 26: Cantidad de Matches - ORB

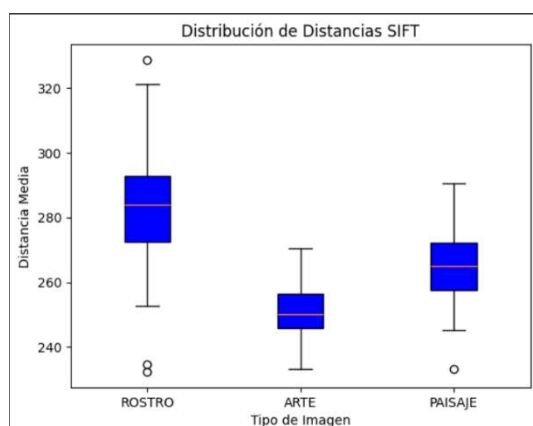


Figura 25: Distancia entre las imágenes - SIFT

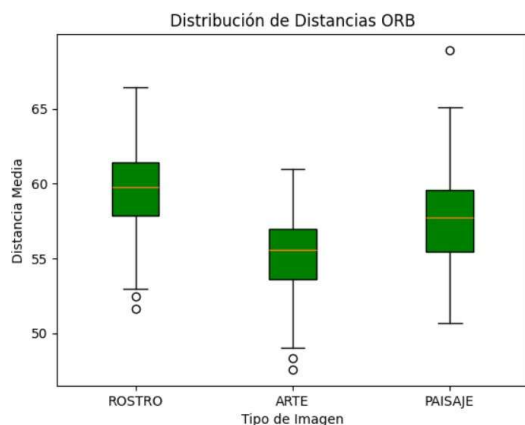


Figura 27: Distancia entre imágenes - ORB

## Stable Diffusion

Métrica	Valor	Desviación Estándar
Distancia máxima	308.92	17.83
Distancia mínima	224.89	-
Distancia media	264.70	14.89
Máximo matches	797	2.87
Mínimo matches	56	-
Promedio de matches	369.68	157.62
Tiempo total	174.65s	
Promedio de características	1609.41	641.75

### ■ ORB

Métrica	Valor	Desviación Estándar
Distancia máxima	66.93	-
Distancia mínima	49.15	-
Distancia media	57.92	3.65
Máximo matches	173	
Mínimo matches	85	-
Promedio de matches	129.46	
Tiempo mínimo	0.1328s	
Tiempo máximo	1.9451s	
Promedio de características	498.88	0.1480

## DIAGRAMA

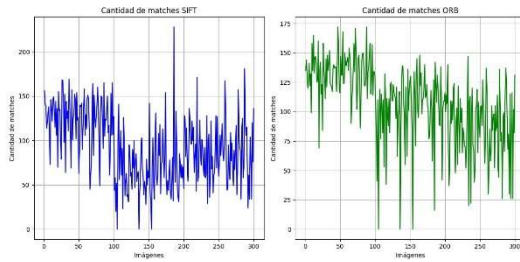


Figura 28: Cantidad de Matches

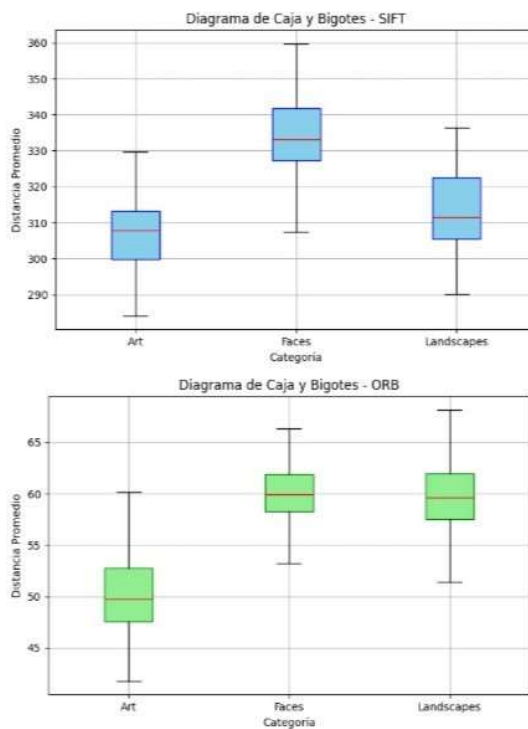


Figura 29: Distancia entre imágenes

Después de analizar los resultados descubrimos que SIFT detecta muchas más características y coincidencias que ORB, aunque el tiempo de procesamiento es más largo. De las tres IA, Disco Diffusion fue la que identificó más detalles (2525 características) y tuvo el mayor número de coincidencias (608), pero también fue la más lenta (378s). Stable Diffusion logró un buen equilibrio, con una cantidad moderada de matches (369) y un tiempo de ejecución menor (175s), mientras que Midjourney detectó menos características y coincidencias (351).

En cuanto a ORB, su ventaja es la velocidad (procesando en milisegundos), pero encontró menos coincidencias en todos los casos.

Midjourney fue el más rápido (0.33s máx), y Disco Diffusion el más preciso en detección de características. En resumen, Disco Diffusion ofrece la mayor riqueza visual y nivel de detalle, Stable Diffusion es el más equilibrado entre precisión y eficiencia, y Midjourney es la mejor opción si lo que buscas es rapidez en el análisis.

## 3. Conclusión

El análisis de similitud entre imágenes generadas por Midjourney, Stable Diffusion y Disco Diffusion con los algoritmos SIFT y ORB mostró diferencias claras en nivel de detalle, precisión y velocidad. Disco Diffusion sobresale al capturar la mayor cantidad de detalles y coincidencias con SIFT (2525 características y 608 matches), aunque esto viene con el costo de un tiempo de procesamiento más alto (378s). Stable Diffusion ofrece un equilibrio entre precisión y eficiencia, logrando un número moderado de matches (369) en un tiempo menor (175s). Midjourney, por su parte, se destacó por su rapidez (0.33s con ORB), pero detectó menos características y coincidencias (351 matches).



En cuanto a los algoritmos, SIFT resultó ser el más preciso al detectar más características y matches, aunque con tiempos de ejecución más largos, mientras que ORB es mucho más rápido pero menos detallado en la detección de similitudes. En definitiva, Disco Diffusion es ideal para quienes buscan la mayor cantidad de detalles, Stable Diffusion representa un balance entre calidad y velocidad, y Midjourney es la mejor opción cuando se necesita rapidez en el análisis de imágenes.

## 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.