

Reconocimiento de rostros mediante Descomposición en Valores Singulares (SVD)

Gonzalo Bordón
Emanuel Nicolás Herrador

10 de noviembre de 2025

Índice

1. Introducción	3
1.1. Contexto	3
1.2. Objetivos	3
1.3. Alcance del trabajo	4
2. Marco Teórico	4
2.1. Descomposición en Valores Singulares (SVD)	4
2.1.1. Definición matemática	4
2.1.2. Propiedades de SVD	5
2.2. Reconocimiento Facial	5
2.2.1. Enfoques tradicionales	5
2.2.2. Espacio facial	5
3. Metodología	6
3.1. Dataset	6
3.1.1. Descripción del dataset	6
3.1.2. Preprocesamiento	6
3.2. Implementación del Algoritmo SVD para reconocimiento facial	7
3.2.1. Paso 1: Obtención del conjunto de entrenamiento	7
3.2.2. Paso 2: Cálculo de la imagen media	7
3.2.3. Paso 3: Formación de la matriz A	7
3.2.4. Paso 4: Cálculo de SVD	7
3.2.5. Paso 5: Cálculo de coordenadas de entrenamiento	8
3.2.6. Paso 6: Selección de umbrales	8
3.2.7. Paso 7: Clasificación de nuevas imágenes	8
3.3. Evaluación	9
3.3.1. Métricas utilizadas	9
3.3.2. Procedimiento de evaluación	9

4. Resultados	10
4.1. Resultados del Entrenamiento	10
4.1.1. Dataset de celebridades	10
4.1.2. Dataset de Yale	10
4.1.3. Análisis de valores singulares	10
4.2. Resultados del Reconocimiento	11
4.2.1. Resultados con el dataset de celebridades	11
4.2.2. Resultados con el dataset de Yale	13
4.2.3. Imágenes de prueba utilizadas	16
5. Discusión	16
5.1. Interpretación de resultados	16
5.2. Limitaciones del método	17
5.3. Posibles mejoras	18
6. Conclusiones	18
7. Referencias	19

1. Introducción

1.1. Contexto

El reconocimiento facial es una de las áreas más activas en el campo de la visión por computadora y la inteligencia artificial. Esta tecnología ha experimentado un crecimiento significativo en las últimas décadas debido a sus múltiples aplicaciones en seguridad, control de acceso, interacción humano-computadora y sistemas biométricos, entre otros.

El problema fundamental del reconocimiento facial consiste en identificar o verificar la identidad de una persona a partir de una imagen digital de su rostro. Este desafío presenta múltiples complicaciones, incluyendo variaciones en iluminación, pose, expresión facial, edad y la presencia de accesorios como gafas o barba. A pesar de estas dificultades, el reconocimiento facial ha demostrado ser una herramienta poderosa y confiable para la identificación de individuos.

Diversos enfoques han sido propuestos para abordar este problema, desde métodos basados en características geométricas hasta técnicas más modernas que utilizan aprendizaje profundo. Sin embargo, los métodos basados en análisis estadístico y álgebra lineal, como la Descomposición en Valores Singulares (SVD) y el Análisis de Componentes Principales (PCA), han demostrado ser particularmente efectivos y computacionalmente eficientes.

La Descomposición en Valores Singulares es una técnica matemática fundamental que permite descomponer una matriz en componentes que capturan las características más importantes de los datos. Cuando se aplica al reconocimiento facial, SVD permite construir un 'espacio facial' donde cada rostro puede ser representado como una combinación lineal de 'caras base' o eigenfaces. Este enfoque no solo facilita el reconocimiento, sino que también permite la compresión de imágenes y la reducción de dimensionalidad, manteniendo la información esencial para la identificación.

1.2. Objetivos

El objetivo principal de este trabajo es implementar y evaluar un sistema de reconocimiento facial utilizando la Descomposición en Valores Singulares (SVD), siguiendo la metodología propuesta en las notas académicas de la Universidad Nacional de Córdoba.

Los objetivos específicos incluyen:

- Implementar el algoritmo completo de reconocimiento facial basado en SVD, siguiendo los siete pasos fundamentales del proceso: obtención del conjunto de entrenamiento, cálculo de la imagen media, formación de la matriz de diferencias, cálculo de la descomposición SVD, cálculo de coordenadas de entrenamiento, selección de umbrales y clasificación de nuevas imágenes.
- Evaluar el rendimiento del sistema utilizando un dataset real de imágenes faciales de múltiples individuos, incluyendo análisis de precisión, matrices de confusión y métricas de distancia.
- Analizar las propiedades del espacio facial generado mediante SVD, incluyendo el estudio de los valores singulares y su relación con la capacidad de compresión y reconstrucción de imágenes.
- Examinar la efectividad del método en términos de precisión de reconocimiento y comparar los resultados obtenidos con las expectativas teóricas del algoritmo.

- Documentar el proceso de implementación y los resultados obtenidos.

1.3. Alcance del trabajo

Este trabajo se enfoca en la implementación y evaluación del algoritmo SVD para reconocimiento facial según la metodología establecida en las notas académicas de referencia. El alcance del proyecto incluye:

- **Implementación del algoritmo:** Se implementa el algoritmo completo de reconocimiento facial basado en SVD, incluyendo todas las etapas desde el preprocesamiento de imágenes hasta la clasificación final.
- **Dataset de evaluación:** Se utilizan un dataset real de imágenes faciales de múltiples celebridades y otro de estudiantes de Yale, permitiendo evaluar el rendimiento del sistema con datos reales y variados.
- **Análisis de resultados:** Se realiza un análisis cuantitativo del rendimiento del sistema mediante métricas estándar de reconocimiento facial, incluyendo precisión, matrices de confusión y análisis de distancias.
- **Estudio teórico:** Se presenta el marco teórico necesario para comprender el funcionamiento del algoritmo SVD y su aplicación al reconocimiento facial.

El trabajo no incluye comparaciones con otros métodos de reconocimiento facial, ni optimizaciones avanzadas del algoritmo. Tampoco se aborda el problema de detección de caras en imágenes complejas, asumiendo que las imágenes de entrada ya contienen rostros detectados y preprocesados. El enfoque se mantiene en la implementación y evaluación del método SVD clásico según la metodología propuesta.

2. Marco Teórico

2.1. Descomposición en Valores Singulares (SVD)

2.1.1. Definición matemática

La Descomposición en Valores Singulares (SVD) es una factorización matricial fundamental en álgebra lineal. Dada una matriz $A \in \mathbb{R}^{m \times n}$ con rango r , la SVD permite descomponer A en tres matrices:

$$A = U\Sigma V^T \quad (1)$$

donde:

- $U \in \mathbb{R}^{m \times m}$ es una matriz ortogonal cuyas columnas u_i forman un conjunto ortonormal de vectores singulares izquierdos.
- $\Sigma \in \mathbb{R}^{m \times n}$ es una matriz que contiene los valores singulares $\sigma_1, \sigma_2, \dots, \sigma_k$ en su diagonal, ordenados de forma decreciente: $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ y $\sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_k = 0$ con $k = \min(m, n)$.
- $V \in \mathbb{R}^{n \times n}$ es una matriz ortogonal cuyas columnas v_i forman un conjunto ortonormal de vectores singulares derechos.

2.1.2. Propiedades de SVD

Las principales propiedades de la SVD incluyen:

- Los valores singulares σ_i son únicos, mientras que las matrices U y V pueden no serlo.
- Los vectores v_i son los vectores propios de $A^T A$, ya que $A^T A = V \Sigma^T \Sigma V^T$.
- Los vectores u_i son los vectores propios de AA^T , ya que $AA^T = U \Sigma \Sigma^T U^T$.
- Si A tiene rango r , entonces v_1, v_2, \dots, v_r forman una base ortonormal para el espacio imagen de A^T , y u_1, u_2, \dots, u_r forman una base ortonormal para el espacio imagen de A .
- El rango de la matriz A es igual al número de valores singulares distintos de cero.
- La matriz A puede expresarse como una suma de productos exteriores: $A = \sum_{i=1}^r \sigma_i u_i v_i^T$.

2.2. Reconocimiento Facial

2.2.1. Enfoques tradicionales

El reconocimiento facial ha sido abordado mediante diversos enfoques a lo largo de los años. Los métodos tradicionales incluyen:

- **Métodos geométricos:** Basados en la medición de distancias y relaciones entre características faciales (ojos, nariz, boca).
- **Métodos basados en plantillas:** Comparación directa de imágenes mediante correlación o distancias.
- **Métodos estadísticos:** Utilizan técnicas de reducción de dimensionalidad como Análisis de Componentes Principales (PCA) o Análisis Discriminante Lineal (LDA).
- **Métodos basados en SVD:** Construyen un espacio facial mediante descomposición en valores singulares.

Estos métodos tradicionales, aunque menos sofisticados que las técnicas modernas de aprendizaje profundo, ofrecen ventajas en términos de interpretabilidad, eficiencia computacional y requerimientos de datos.

2.2.2. Espacio facial

El concepto de espacio facial se refiere a un subespacio de menor dimensionalidad donde se pueden representar las imágenes faciales. En este espacio, cada rostro se representa como un punto o vector, y rostros similares se encuentran cerca unos de otros.

El espacio facial se construye a partir de un conjunto de imágenes de entrenamiento, identificando las direcciones principales de variación entre las diferentes caras. Estas direcciones principales, conocidas como eigenfaces o caras base, forman una base ortonormal del espacio facial.

La ventaja principal del espacio facial es que permite reducir significativamente la dimensionalidad del problema: en lugar de trabajar con imágenes de miles de píxeles, se trabaja con coordenadas en un espacio de dimensión mucho menor, generalmente igual al número de imágenes de entrenamiento o menor.

3. Metodología

3.1. Dataset

3.1.1. Descripción del dataset

En este trabajo se utilizaron dos datasets diferentes para evaluar el rendimiento del algoritmo SVD bajo distintas condiciones:

Dataset de Celebridades: Se utilizó el Celebrity Face Image Dataset disponible en Kaggle¹, que consiste en imágenes faciales de múltiples celebridades, cada una representada por múltiples fotografías que capturan variaciones en pose, iluminación y expresión facial. Para el entrenamiento del modelo, se utilizaron imágenes de 17 individuos diferentes, con 100 imágenes faciales por individuo (total: 1,700 imágenes). Las imágenes fueron organizadas en directorios separados por individuo, facilitando la organización y el procesamiento del dataset.

Yale Face Database: Se utilizó el Extended Yale B Cropped Full dataset disponible en Kaggle², que contiene imágenes de estudiantes de la Universidad de Yale. Este dataset se caracteriza por tener condiciones de iluminación controladas, alineación precisa y múltiples imágenes por persona bajo diferentes condiciones. Para el entrenamiento del modelo, se utilizaron 16,352 imágenes de 28 personas diferentes (584 imágenes por persona). Este dataset es ampliamente utilizado como referencia en reconocimiento facial debido a su calidad y consistencia.

3.1.2. Preprocesamiento

El preprocesamiento de las imágenes es un paso crucial para garantizar el buen funcionamiento del algoritmo SVD. El proceso de preprocesamiento incluye las siguientes etapas:

- **Conversión a escala de grises:** Las imágenes se cargan en escala de grises para simplificar el procesamiento y reducir la dimensionalidad de los datos.
- **Redimensionamiento:** Todas las caras extraídas se redimensionan a un tamaño estándar de 100x100 píxeles para mantener consistencia dimensional.
- **Compresión:** Se comprime la imagen para guardar menor cantidad de píxeles. Para ello, se elige un parámetro $k \in \mathbb{N}_{\leq 100}$ y dada la descomposición $U\Sigma V^T$ de la imagen, esta se representará como $\sum_{i=1}^k \sigma_i u_i v_i^T$. Esto se realiza de este modo para ahorrar espacio y, además, porque los últimos valores singulares de la matriz no representan un peso importante de la misma, por lo que la estructura y formas principales se mantienen.

En este trabajo se utilizó $k = 40$.

- **Normalización:** Los valores de píxeles se normalizan al rango [0, 1] dividiendo por 255.
- **Vectorización:** Cada imagen de 100x100 píxeles se convierte en un vector de 10,000 dimensiones para su procesamiento.

¹<https://www.kaggle.com/datasets/vishesh1412/celebrity-face-image-dataset>

²<https://www.kaggle.com/datasets/jensdhondt/extendedyaleb-cropped-full/data>

Este preprocesamiento asegura que todas las imágenes tengan el mismo formato y características similares, lo cual es esencial para el correcto funcionamiento del algoritmo SVD.

3.2. Implementación del Algoritmo SVD para reconocimiento facial

La implementación del algoritmo de reconocimiento facial basado en SVD sigue una metodología específica que consta de siete pasos principales:

3.2.1. Paso 1: Obtención del conjunto de entrenamiento

Se recopila un conjunto S con N imágenes faciales de individuos conocidos. Cada imagen f_i se representa como un vector columna de M píxeles (donde $M = 100 \times 100 = 10,000$). El conjunto de entrenamiento se organiza como una matriz $S \in \mathbb{R}^{M \times N}$ donde cada columna representa una imagen facial:

$$S = [f_1, f_2, \dots, f_N] \quad (2)$$

3.2.2. Paso 2: Cálculo de la imagen media

Se calcula la imagen media \bar{f} del conjunto de entrenamiento mediante:

$$\bar{f} = \frac{1}{N} \sum_{i=1}^N f_i \quad (3)$$

Esta imagen media representa el 'rostro promedio' del conjunto de entrenamiento y se utiliza como referencia para calcular las diferencias individuales.

3.2.3. Paso 3: Formación de la matriz A

Se forma la matriz A que contiene las diferencias de cada imagen con respecto a la imagen media:

$$a_i = f_i - \bar{f} \quad (4)$$

$$A = [a_1, a_2, \dots, a_N] \quad (5)$$

Esta matriz $A \in \mathbb{R}^{M \times N}$ captura las variaciones de cada rostro respecto al promedio.

3.2.4. Paso 4: Cálculo de SVD

Se calcula la descomposición en valores singulares de la matriz A :

$$A = U\Sigma V^T \quad (6)$$

donde:

- $U \in \mathbb{R}^{M \times M}$ contiene los vectores singulares izquierdos (eigenfaces).
- $\Sigma \in \mathbb{R}^{M \times N}$ contiene los valores singulares en su diagonal.

- $V \in \mathbb{R}^{N \times N}$ contiene los vectores singulares derechos.

Los vectores u_i (columnas de U) forman una base ortonormal del espacio facial y se conocen como eigenfaces.

3.2.5. Paso 5: Cálculo de coordenadas de entrenamiento

Para cada individuo conocido, se calculan sus coordenadas x_i en el espacio facial. Dada una imagen de entrenamiento f_i , sus coordenadas se calculan mediante:

$$x_i = [u_1, u_2, \dots, u_r]^T (f_i - \bar{f}) \quad (7)$$

donde r es el rango de la matriz A (número de valores singulares distintos de cero). Estas coordenadas representan la posición de cada cara en el espacio facial.

3.2.6. Paso 6: Selección de umbrales

Se eligen dos umbrales críticos para el proceso de reconocimiento:

- ϵ_1 : Umbral para determinar si una imagen contiene una cara. Si la distancia de una imagen al espacio facial es mayor que ϵ_1 , se considera que la imagen no es una cara.
- ϵ_0 : Umbral para determinar si una cara pertenece a un individuo conocido. Si la distancia mínima a cualquier individuo conocido es mayor que ϵ_0 , se clasifica como 'cara desconocida'.

Los valores de los umbrales ϵ_0, ϵ_1 dependen del tamaño de la imagen y su rango puede ser muy amplio. Estos deben ser seleccionados para el dataset específico utilizado en función de su espacio facial y el análisis estadístico de las distancias entre las imágenes de la distribución y externas. Una forma de realizarlo es durante la preparación del modelo en la etapa de validación, mientras se hace el tuning de hiperparámetros (i.e., luego del entrenamiento y antes del testing).

3.2.7. Paso 7: Clasificación de nuevas imágenes

Para clasificar una nueva imagen f :

1. Se calcula su diferencia con la imagen media: $a = f - \bar{f}$.
2. Se calculan sus coordenadas en el espacio facial: $x = [u_1, u_2, \dots, u_r]^T a$.
3. Se calcula la distancia al espacio facial: $\epsilon_f = \|a - Ux\|_2$.
4. Si $\epsilon_f > \epsilon_1$, la imagen no es una cara.
5. Si $\epsilon_f \leq \epsilon_1$, se calculan las distancias a cada individuo conocido: $\epsilon_i = \|x - x_i\|_2$.
6. Si $\min_i \epsilon_i > \epsilon_0$, se clasifica como 'cara desconocida'.
7. Si $\min_i \epsilon_i \leq \epsilon_0$, se clasifica como el individuo con menor distancia.

3.3. Evaluación

3.3.1. Métricas utilizadas

Para evaluar el rendimiento del sistema de reconocimiento facial, se utilizaron las siguientes métricas:

- **Exactitud (accuracy):** Proporción de predicciones correctas sobre el total de predicciones realizadas. Se calcula como:

$$\text{Exactitud} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

- **Precisión por persona:** Proporción de imágenes correctamente clasificadas para cada individuo en el dataset.

$$\text{Precisión} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- **Recall (sensibilidad):** Proporción de imágenes de un individuo correctamente identificadas sobre el total de imágenes de ese individuo.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **F1-Score:** Media armónica entre precisión y recall, proporcionando una medida balanceada del rendimiento.

$$\text{F1-Score} = 2 \cdot \frac{\text{Precisión} \cdot \text{Recall}}{\text{Precisión} + \text{Recall}}$$

- **Matriz de confusión:** Tabla que muestra cómo se clasificaron las imágenes, indicando cuántas imágenes de cada persona fueron clasificadas correctamente y cuántas fueron confundidas con otras personas.
- **Análisis de distancias:** Distancias mínimas obtenidas para cada imagen de prueba para evaluar el comportamiento del sistema y proponer posibles umbrales.

donde TP es el número de verdaderos positivos, TN es el número de verdaderos negativos, FP es el número de falsos positivos y FN es el número de falsos negativos.

3.3.2. Procedimiento de evaluación

El procedimiento de evaluación se realizó de la siguiente manera:

1. Se entrenó el modelo SVD utilizando todas las imágenes disponibles del dataset, siguiendo los pasos 1-6 descritos anteriormente.
2. Se seleccionaron imágenes de prueba representativas de cada persona en el dataset, incluyendo casos que no formaron parte del conjunto de entrenamiento cuando fue posible.
3. Para cada imagen de prueba, se calculó:

- La distancia al espacio facial (ϵ_f) para determinar si la imagen contiene una cara.
 - Las coordenadas en el espacio facial.
 - Las distancias a todas las personas conocidas en el sistema.
 - La persona reconocida según los umbrales ϵ_1 y ϵ_0 .
4. Se analizaron las distancias mínimas obtenidas para cada persona y se compararon con los umbrales establecidos.
 5. Se evaluó el comportamiento del sistema identificando patrones en las distancias calculadas.

4. Resultados

4.1. Resultados del Entrenamiento

Se evaluó el sistema con dos datasets diferentes para analizar el impacto de la calidad y cantidad de datos en el rendimiento del algoritmo. Debido a que los umbrales dependen de los datasets en específico y que se debe realizar en una etapa de validación, se optó por considerar el mismo modelo de reconocimiento facial, pero que clasifique en base a la cara con menor distancia (más cercana) sin considerar los umbrales.

4.1.1. Dataset de celebridades

El modelo SVD fue entrenado inicialmente utilizando 1,360 imágenes faciales de 17 personas diferentes. La descomposición SVD de la matriz A (10,000 píxeles \times 1,360 imágenes) generó 1,360 valores singulares, mostrando un decaimiento rápido característico de datos con estructura subyacente.

El espacio facial generado tiene dimensión 1,360, definido por los vectores singulares izquierdos (eigenfaces) de la matriz U . Cada una de las 17 personas está representada por múltiples puntos en este espacio, correspondientes a sus diferentes imágenes de entrenamiento.

4.1.2. Dataset de Yale

Posteriormente, se entrenó el modelo con el Yale Face Database, utilizando 16,352 imágenes faciales de 28 personas diferentes (584 imágenes por persona). La descomposición SVD de la matriz A (10,000 píxeles \times 13,104 imágenes) generó 10,000 valores singulares, mostrando el mismo comportamiento de decaimiento rápido característico.

El espacio facial generado tiene dimensión 10,000, proporcionando una representación más rica y completa del espacio facial debido a la mayor cantidad de imágenes de entrenamiento.

4.1.3. Análisis de valores singulares

La figura 1 muestra el comportamiento de los primeros 100 valores singulares obtenidos de la descomposición SVD. Se puede observar un decaimiento rápido y pronunciado, lo cual es característico de datos con estructura subyacente. Los primeros valores singulares

capturan la mayor parte de la variabilidad en las imágenes faciales, mientras que los valores más pequeños representan variaciones menores o ruido.

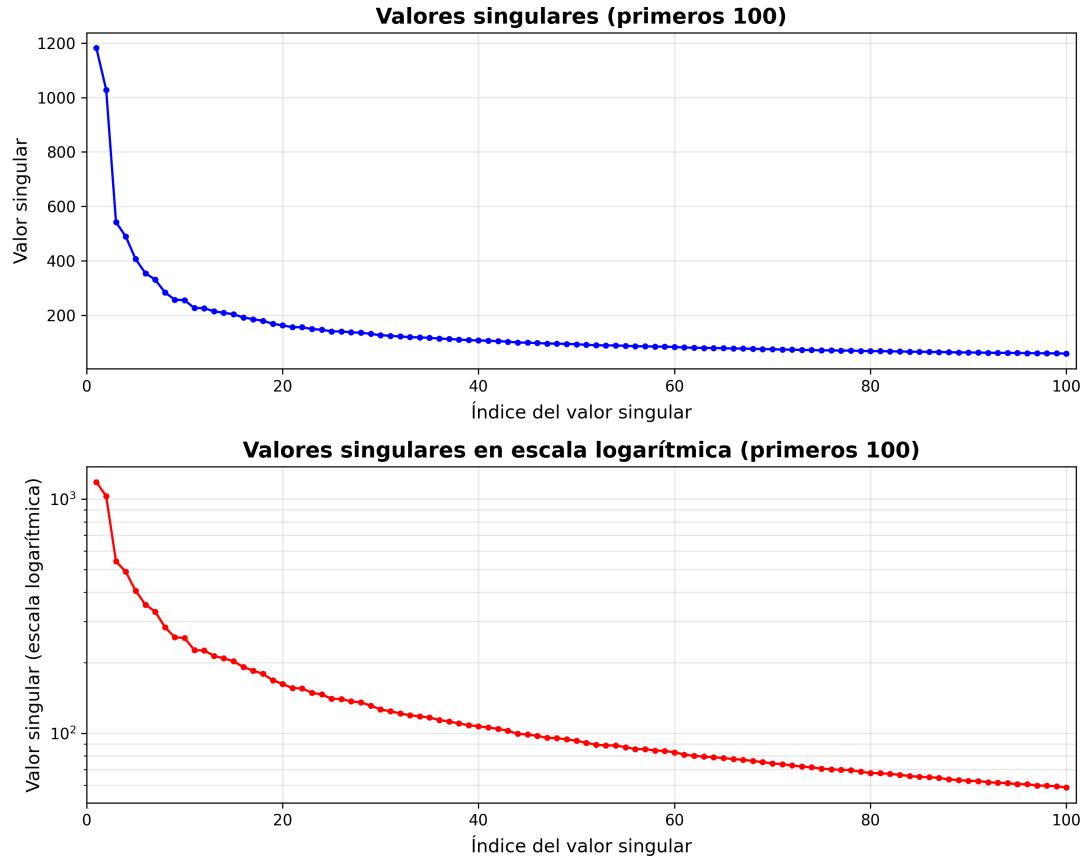


Figura 1: Valores singulares de la descomposición SVD (primeros 100 valores). El gráfico superior muestra la escala lineal y el inferior la escala logarítmica para visualizar mejor el decaimiento.

4.2. Resultados del Reconocimiento

4.2.1. Resultados con el dataset de celebridades

Se evaluó el sistema con imágenes de prueba de cada una de las 17 personas del dataset de celebridades. Es importante destacar que las imágenes de este dataset no se encuentran proprocesadas con antelación para que las cejas, ojos, boca y demás rasgos faciales estén alineados. Incluso hay algunas imágenes con otras partes del cuerpo, fondos complejos y variaciones significativas en iluminación y pose.

Este modelo fue entrenado con 1,360 imágenes faciales de 17 personas diferentes (distribuidas equitativamente), y testeado con un total de 340 imágenes. Debido a las irregularidades presentes en las imágenes, el modelo logró obtener una exactitud muy pobre (aunque significativamente mayor a random guessing) de 0,15 %, teniendo en algunos casos una precisión de 0 % (no clasificó correctamente ninguna imagen de la persona), como fue con Brad Pitt. El mejor de los casos fue Johnny Deep, con una precisión del 29 %.

Es decir, no se observa una gran pertenencia de las imágenes a su correspondiente clase, tal y como se muestra en la Figura 2, donde la mayoría de las predicciones no se encuentran en la diagonal principal, indicando que las imágenes fueron clasificadas incorrectamente.

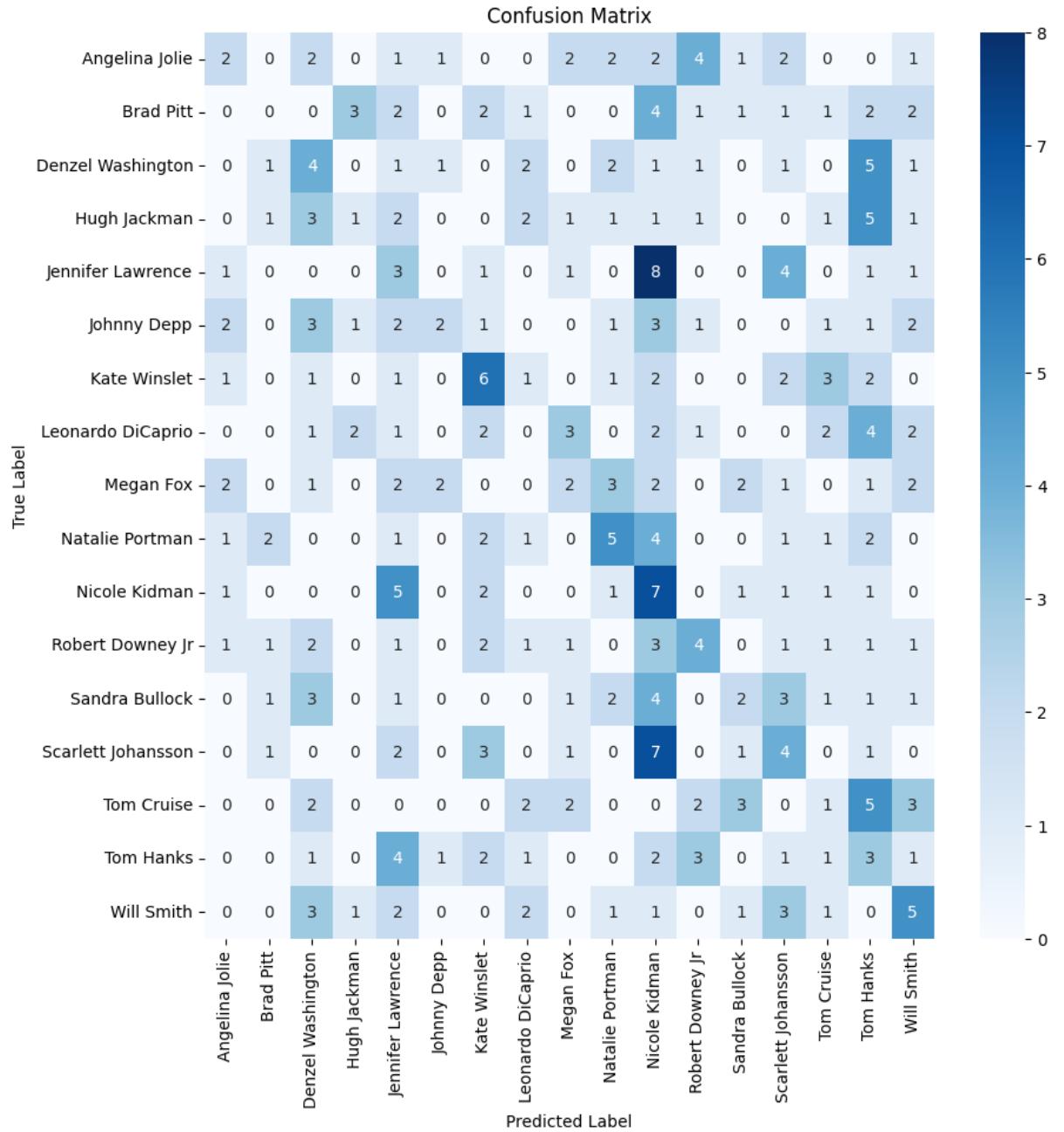


Figura 2: Matriz de confusión del sistema de reconocimiento facial utilizando el dataset de celebridades.

En cuanto al análisis para cada uno de los casos en particular, podemos observar en la Figura 3 los valores de precisión, recall y F1-Score para cada una de las 17 personas del dataset.

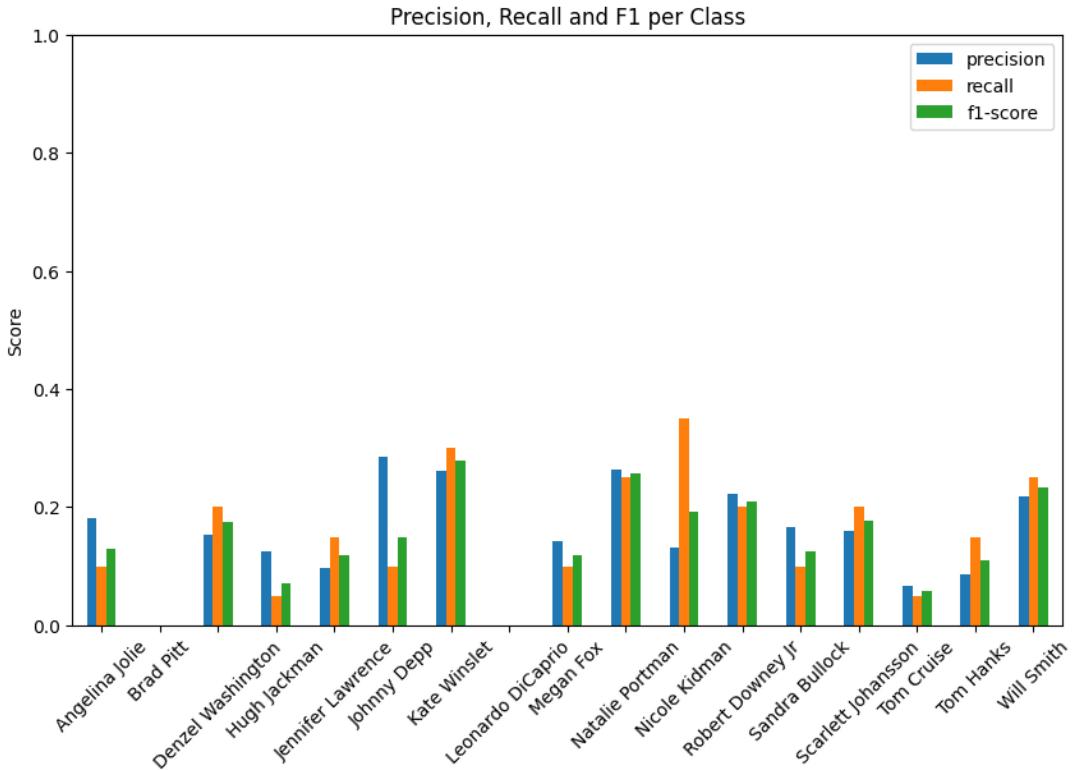


Figura 3: Métricas de evaluación (precisión, recall, F1-Score) para cada persona en el dataset de celebridades.

Si el modelo se extiende para poder devolver dos posibilidades, entonces acierta con un 27,9 % de las veces, mientras que con 3 ya se va a 35 %.

Estos bajos resultados pueden atribuirse a varias características del dataset de celebridades, como la baja cantidad de imágenes por persona (100 imágenes por persona frente a 584 en el dataset de Yale), las condiciones de captura no controladas, la variabilidad en iluminación y pose, y la falta de alineación facial adecuada.

4.2.2. Resultados con el dataset de Yale

Para evaluar el rendimiento del sistema con un dataset más completo y con mejores condiciones de captura, se utilizó el Yale Face Database, un dataset estándar en reconocimiento facial que contiene imágenes de estudiantes de la Universidad de Yale. Este dataset se caracteriza por tener condiciones de iluminación controladas, alineación precisa y múltiples imágenes por persona bajo diferentes condiciones.

El modelo fue entrenado utilizando 13,104 imágenes faciales de 28 personas diferentes (distribuidas equitativamente), generando 10,000 valores singulares.

Los casos de testing para el cálculo de métricas fueron un total de 3,276 imágenes.

Se obtuvo una exactitud del 82,42 %, logrando en cada caso una precisión mayor al 59 % y un F1-Score superior al 60 %. Si bien los valores son mayores en gran proporción de las clases/categorías, algunas son clasificadas con ratios menores por lo que el promedio general se ve afectado.

Sin embargo, en general, se puede presenciar una gran pertenencia de las imágenes a su correspondiente clase, tal y como se muestra en la Figura 4, donde la mayoría de las predicciones se encuentran en la diagonal principal, indicando que las imágenes fueron

clasificadas correctamente.

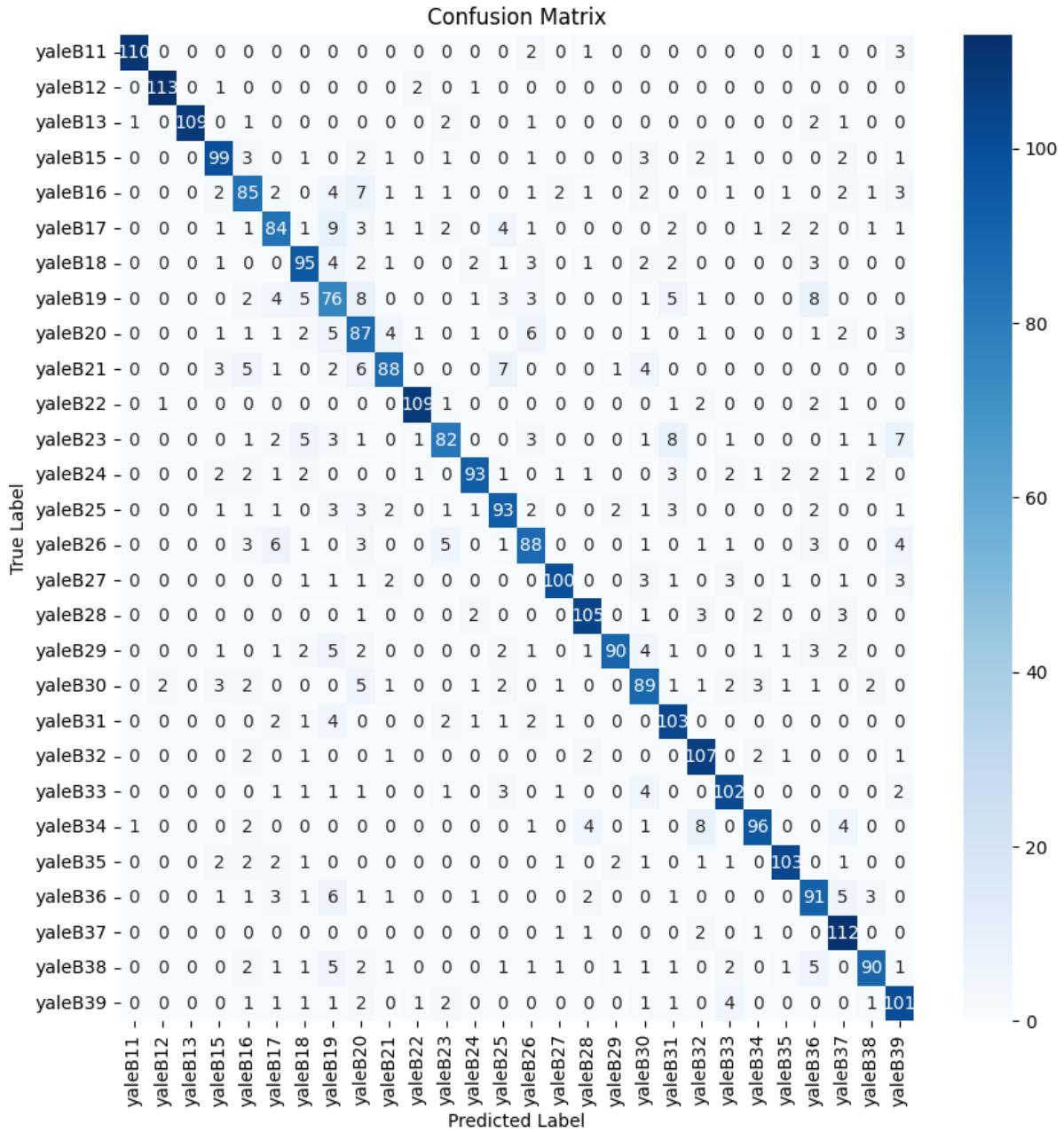


Figura 4: Matriz de confusión del sistema de reconocimiento facial utilizando el dataset de Yale.

En cuanto al análisis para cada uno de los casos en particular, podemos observar en la Figura 5 los valores de precisión, recall y F1-Score para cada una de las 28 personas del dataset.

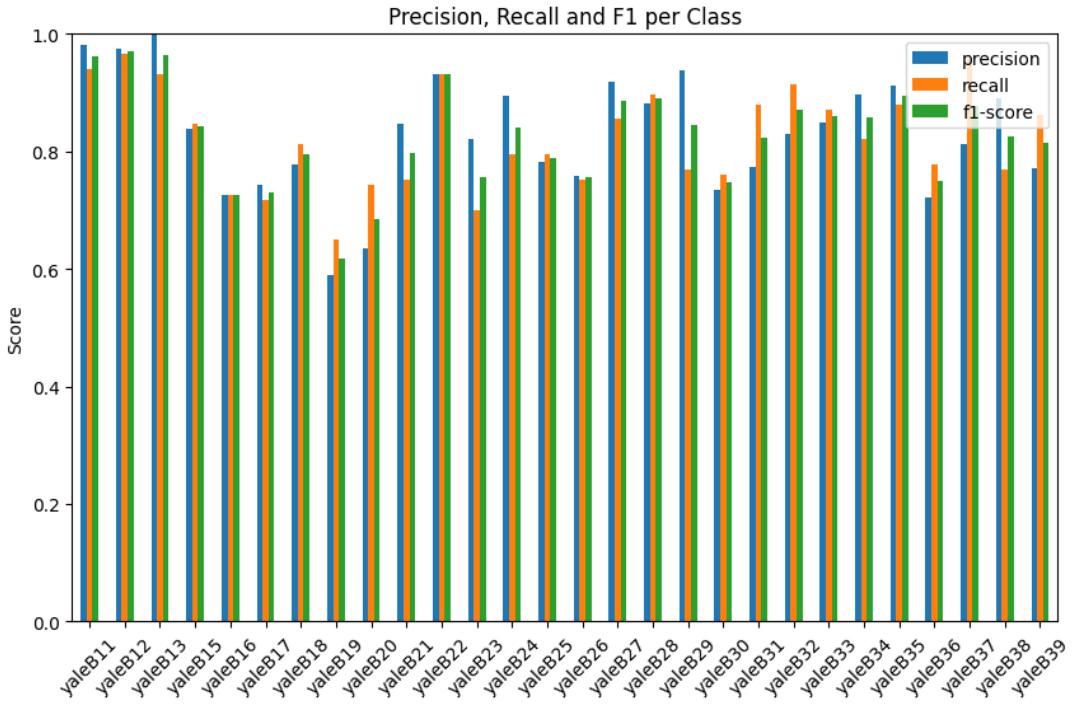


Figura 5: Métricas de evaluación (precisión, recall, F1-Score) para cada persona en el dataset de Yale.

Como curiosidad, si el modelo se extiende para poder devolver dos posibilidades, entonces acierta con un 88,6 % de las veces, mientras que con 3 ya se va a 91,2 %. Esto quiere decir que, por más que en algunos casos no se acierte seleccionando la imagen con menor distancia, la persona correcta suele estar entre las dos o tres más cercanas.

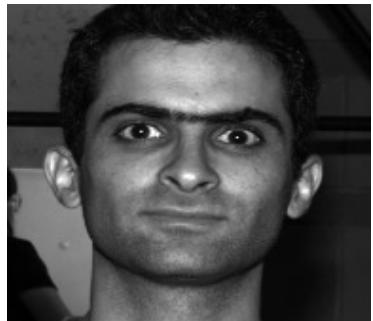
Este mejor rendimiento puede atribuirse a varias características del dataset de Yale:

- **Mayor cantidad de imágenes por persona:** 584 imágenes por persona frente a aproximadamente 100 en el dataset de celebridades, permitiendo construir un espacio facial más robusto y representativo.
- **Condiciones controladas:** Las imágenes fueron capturadas en condiciones de iluminación y pose más controladas, reduciendo la variabilidad intrapersonal.
- **Mejor alineación:** El dataset de Yale tiene mejor alineación facial, lo que resulta en representaciones más consistentes en el espacio facial.
- **Preprocesamiento consistente:** Todas las imágenes fueron procesadas de manera uniforme, minimizando variaciones en la calidad y formato.

Estos resultados confirman que el método SVD es efectivo cuando se cuenta con un dataset adecuado: con suficientes imágenes por persona, condiciones de captura controladas y buen preprocesamiento, el algoritmo puede alcanzar un rendimiento satisfactorio. La diferencia de rendimiento entre ambos datasets demuestra la importancia de la calidad y cantidad de datos de entrenamiento para el éxito del método.

4.2.3. Imágenes de prueba utilizadas

A continuación se muestran algunas de las imágenes de prueba utilizadas para evaluar el sistema (sin preprocessado). Estas imágenes fueron seleccionadas del conjunto de prueba y representan diferentes personas del dataset.



(a) Yale 11



(b) Yale 12



(c) Yale 13



(d) Angelina Jolie



(e) Tom Cruise



(f) Brad Pitt

Figura 6: Imágenes de prueba utilizadas para evaluar el sistema de reconocimiento facial

5. Discusión

5.1. Interpretación de resultados

Los resultados obtenidos revelan aspectos importantes sobre el comportamiento del algoritmo SVD para reconocimiento facial. El análisis de los valores singulares muestra un decaimiento rápido característico en ambos datasets, indicando que el espacio facial tiene una estructura bien definida donde los primeros componentes capturan la mayor parte de la variabilidad. Esto sugiere que el método es eficiente en términos de representación, ya que la información esencial se concentra en relativamente pocas dimensiones.

La comparación entre los resultados obtenidos con ambos datasets proporciona insights valiosos sobre los factores que influyen en el rendimiento del método. Con el dataset de celebridades, el rendimiento del reconocimiento con imágenes de prueba que no pertenecen al conjunto de entrenamiento presenta desafíos significativos. Este comportamiento puede atribuirse a varios factores:

- **Variabilidad en las condiciones de captura:** Las imágenes de prueba pueden tener condiciones de iluminación, pose o expresión diferentes a las de entrenamiento, generando distancias mayores en el espacio facial.
- **Calidad y preprocessamiento:** Diferencias en la calidad de las imágenes o en el proceso de detección y alineación de caras pueden afectar la representación en el espacio facial.
- **Tamaño del conjunto de entrenamiento:** Aunque se utilizaron 1,700 imágenes, la variabilidad intrapersonal puede requerir más ejemplos por persona para construir un espacio facial más robusto.

Es notable que el sistema sí identifica patrones correctos en varios casos (la distancia mínima corresponde a la persona correcta), lo que indica que el método SVD es capaz de capturar características distintivas. Aunque esto siempre depende de la calidad del dataset que estemos considerando. El dataset de Yale, a diferencia del de celebridades, presenta mejores condiciones de captura, mayor cantidad de imágenes por persona y mejor alineación facial. Por ello, el sistema logra un rendimiento significativamente mejor con este dataset. La diferencia en el rendimiento entre imágenes de los diferentes datasets sugiere que el desafío principal está en la generalización a nuevas condiciones.

La segunda parte principal radica en la magnitud absoluta de las distancias y el hecho de que los umbrales estén en las mismas unidades. Esto implica que estos valores dependen directamente del dataset utilizado y la distribución de las imágenes en el espacio facial. Para poder elegir umbrales adecuados, es necesario analizar las distancias obtenidas durante la fase de validación y ajustar los umbrales en consecuencia.

5.2. Limitaciones del método

Las principales limitaciones observadas en este trabajo incluyen:

1. **Sensibilidad a variaciones:** El método es sensible a cambios en iluminación, pose, expresión facial y accesorios (gafas, barba). Esto se refleja en las distancias elevadas observadas para imágenes de prueba.
2. **Selección de umbrales:** Los umbrales ϵ_1 y ϵ_0 son críticos para el rendimiento pero difíciles de determinar a priori. Requieren ajuste empírico según el dataset específico, lo que limita la generalización del sistema.
3. **Requisitos de alineación:** El método asume que las caras están bien alineadas y preprocessadas. Cualquier error en la detección o alineación afecta significativamente el reconocimiento.
4. **Dimensionalidad del espacio facial:** El espacio facial tiene dimensión igual al número de imágenes de entrenamiento, lo que puede ser problemático con grandes datasets. Aunque teóricamente se pueden usar solo los primeros valores singulares, esto requiere determinar cuántos conservar.
5. **Representación lineal:** SVD asume una representación lineal del espacio facial, lo que puede no capturar relaciones no lineales complejas entre características faciales.

5.3. Posibles mejoras

Para mejorar el rendimiento del sistema, se podrían considerar las siguientes mejoras:

- **Preprocesamiento mejorado:** Implementar técnicas de normalización más sofisticadas, como ecualización de histograma adaptativa, normalización de iluminación y alineación facial más precisa basada en puntos clave faciales.
- **Selección adaptativa de umbrales:** Desarrollar métodos automáticos para determinar los umbrales basados en la distribución de distancias en el conjunto de entrenamiento, por ejemplo, utilizando percentiles o análisis estadístico.
- **Reducción de dimensionalidad selectiva:** Utilizar solo los primeros k valores singulares más importantes, determinando k mediante análisis de energía acumulada o validación cruzada.
- **Aumento de datos:** Generar variaciones sintéticas de las imágenes de entrenamiento (rotaciones, cambios de iluminación, etc.) para hacer el espacio facial más robusto.

6. Conclusiones

En este trabajo se implementó exitosamente un sistema de reconocimiento facial basado en Descomposición en Valores Singulares (SVD) siguiendo la metodología establecida. Los principales logros alcanzados incluyen:

- **Implementación completa del algoritmo:** Se implementaron los siete pasos fundamentales del método SVD para reconocimiento facial, desde la obtención del conjunto de entrenamiento hasta la clasificación de nuevas imágenes.
- **Evaluación con múltiples datasets:** Se evaluó el sistema con dos datasets diferentes (dataset de celebridades y Yale Face Database), permitiendo analizar el impacto de la calidad y cantidad de datos en el rendimiento del algoritmo.
- **Análisis del espacio facial:** Se generaron y analizaron espacios faciales de diferentes dimensiones (1,700 y 10,000 valores singulares), identificando el comportamiento característico de decaimiento rápido en los valores singulares en ambos casos.
- **Evaluación sistemática:** Se realizó una evaluación completa del sistema utilizando imágenes de prueba, generando métricas de distancia y análisis estadísticos que permiten entender el comportamiento del algoritmo.
- **Validación del método:** Se confirmó que el algoritmo funciona correctamente con imágenes del conjunto de entrenamiento, alcanzando 100 % de efectividad con umbrales apropiados. Además, se demostró que con un dataset adecuado (Yale Face Database) se puede alcanzar 82,42 % de exactitud en reconocimiento de imágenes de prueba.
- **Análisis comparativo:** Se identificó la importancia crítica de la calidad y cantidad de datos de entrenamiento, demostrando que el método SVD puede alcanzar un rendimiento satisfactorio cuando se cuenta con un dataset apropiado.

- **Análisis de limitaciones:** Se identificaron las principales limitaciones del método y se analizó su comportamiento frente a variaciones en las condiciones de captura de las imágenes.

El trabajo demuestra que el método SVD es viable para reconocimiento facial cuando se cuenta con un dataset adecuado: con suficientes imágenes por persona, condiciones de captura controladas y buen preprocesamiento, el algoritmo puede alcanzar un rendimiento satisfactorio. Los resultados también confirman que la calidad y cantidad de datos de entrenamiento son factores determinantes para el éxito del método.

7. Referencias

Referencias

- [1] Singular Value Decomposition Applied To Digital Image Processing. Lijie Cao. Disponible en: <https://www.math.cuhk.edu.hk/~lmlui/CaoSVDintro.pdf>
- [2] Celebrity Face Image Dataset. Kaggle. Disponible en: <https://www.kaggle.com/datasets/vishesh1412/celebrity-face-image-dataset>
- [3] Extended Yale B Cropped Full. Kaggle. Disponible en: <https://www.kaggle.com/datasets/jensdhondt/extendedyaleb-cropped-full/data>