

Midterm Project

Analizador de Arte Multimedia

Autor: Mateo Cardona Arias

Septiembre 2025

1. Introducción

Este proyecto tiene como objetivo el desarrollo de un analizador multimedia capaz de integrar el procesamiento de imágenes y audio en un mismo flujo de trabajo. El programa permite extraer características visuales y sonoras de archivos de entrada y luego relacionarlas mediante un proceso de fusión multimodal.

Se buscó que la implementación fuese práctica y clara, evitando el uso de funciones encapsuladas, para que la ejecución pueda seguirse de manera lineal. El usuario únicamente debe proporcionar una imagen (.jpg) y un archivo de audio (.wav), y el sistema se encarga de realizar todo el análisis automáticamente.

Además, dentro del proyecto se incluyen varias imágenes de prueba y audios de distintos géneros, como música clásica o rock, de manera que se puedan realizar diferentes pruebas sin necesidad de modificar el código

Aparte se realizó un video explicativo del programa: <https://youtu.be/QURcYFdZRwQ>

2. Tecnologías utilizadas

El proyecto fue implementado en **Python**, utilizando las siguientes bibliotecas:

- **NumPy**: manejo de arreglos y operaciones matemáticas.
- **Matplotlib**: visualización de gráficos y figuras.
- **OpenCV**: procesamiento de imágenes y conversión de espacios de color.
- **Scikit-learn (KMeans)**: extracción de colores dominantes en la imagen.
- **Librosa**: análisis de audio (RMS, centroide espectral, tempo, onset, espectrogramas).

Los resultados se presentan de dos maneras:

1. En consola, donde se listan las principales características extraídas.
2. En la carpeta **outputs/**, que el mismo programa crea automáticamente y donde se almacenan gráficos, espectrogramas y un archivo de texto con una interpretación automática.

3. Descripción del funcionamiento del código

El código está estructurado como un flujo lineal que va ejecutando paso a paso cada etapa del análisis. A continuación, se describe de manera general lo que ocurre durante la ejecución:

1. **Carga de archivos**: se leen la imagen (.jpg) y el audio (.wav) definidos en las variables de configuración. Si los archivos no se encuentran, el programa notifica el error.
2. **Procesamiento de la imagen**: la imagen se convierte del espacio de color BGR a RGB y HSV. Luego:

- Se calculan histogramas de color en RGB y del tono (Hue) en HSV.
 - Se aplica el algoritmo *KMeans* para identificar los colores dominantes.
 - Se obtiene la paleta de colores con su frecuencia de aparición.
3. **Procesamiento del audio:** se carga el archivo de sonido con *Librosa* y, en caso necesario, se eliminan los silencios. Después se extraen características como:
- Energía RMS (intensidad promedio).
 - Centroide espectral (brillo del sonido).
 - Tempo estimado y beats.
 - Tasa de onsets (ataques sonoros).
 - Mel-espectrograma, que representa la energía en distintas bandas de frecuencia a lo largo del tiempo.
4. **Fusión multimodal:** se combinan métricas de imagen (tono, saturación) con métricas de audio (brillo, energía). Esto permite generar un perfil conjunto que refleja la relación entre los colores y las propiedades del sonido.
5. **Visualización y almacenamiento:**
- Se genera una figura con subgráficos: la imagen original, su paleta de colores, histogramas, forma de onda, RMS y centroide espectral.
 - Se guarda un mel-espectrograma en un archivo aparte.
 - Se escribe un archivo de texto con una interpretación automática de los resultados.

4. Resultados

En esta sección se presentan los resultados obtenidos al ejecutar pruebas con los archivos de ejemplo.

En particular, se realizó una prueba utilizando la **imagen “espacio”** y el **audio de “rock”**. Para este caso, se incluyen diferentes tipos de salidas generadas por el programa:

- **Características de la imagen:** gráfica con los colores dominantes, histogramas y paletas obtenidas.
- **Características del audio:** gráfica con el espectro, RMS, centroide espectral, tempo y onsets.
- **Fusión de resultados:** representación conjunta donde se visualizan en un mismo espacio tanto las propiedades de la imagen como del audio.
- **Resultados en consola:** captura de pantalla de los valores numéricos calculados, como los colores dominantes en RGB/HSV y las métricas de audio.

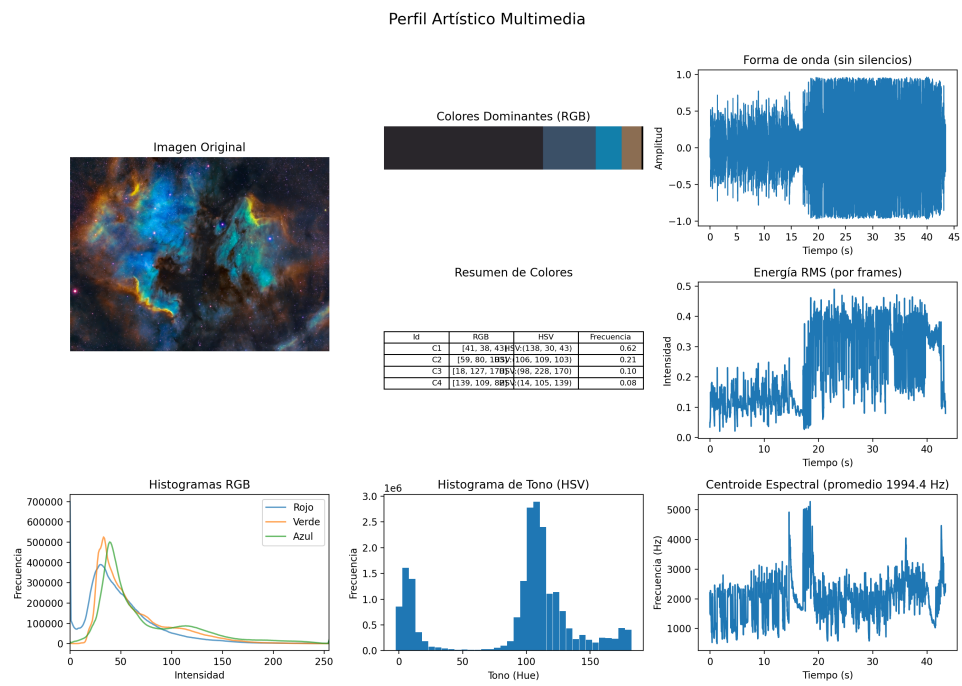


Figura 1: Perfil artístico multimedia.

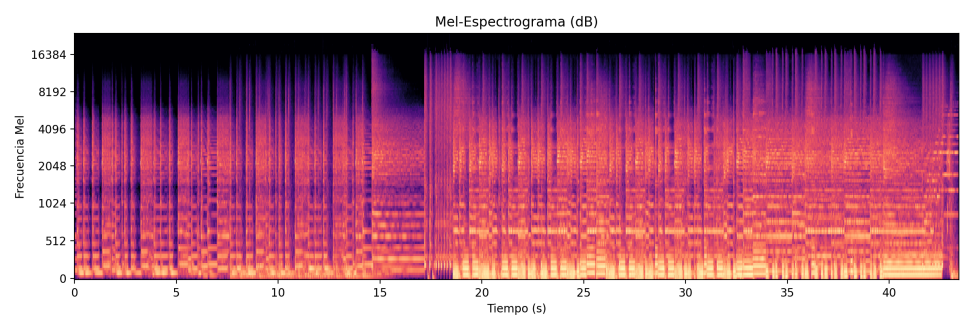


Figura 2: Mel-espectrograma del audio.

```

=== CARACTERÍSTICAS DE LA IMAGEN ===
- Tamaño de la imagen: 5303x3971 píxeles
- Color 1: RGB=[41, 38, 43] | HSV=(138, 30, 43) | Frecuencia=0.61
- Color 2: RGB=[57, 82, 106] | HSV=(105, 118, 106) | Frecuencia=0.20
- Color 3: RGB=[16, 120, 175] | HSV=(90, 232, 175) | Frecuencia=0.09
- Color 4: RGB=[134, 101, 74] | HSV=(14, 114, 134) | Frecuencia=0.09

=== CARACTERÍSTICAS DEL AUDIO ===
- Frecuencia de muestreo: 48000 Hz
- Duración: 45.05s
- Canales: 1
- Energía RMS promedio: 0.2221
- Centroide espectral promedio: 1994.4 Hz
- Tempo estimado: 125.0 BPM
- Tasa de onsets promedio: 1.00

=== FUSIÓN MULTIMODAL ===
- Calidez: 0.51
- Saturación: 0.48
- Brillo: 0.08
- Energía: 0.45
c:\Users\teoca\OneDrive\Escritorio\Universidad\Electiva\MidtermProject\multimedia_art_analyzer.py:200: UserWarning: The figure layout has
changed to tight
  fig.tight_layout(rect=[0, 0, 1, 0.97])

Procesamiento completado. Archivos guardados en 'outputs/'.

```

Figura 3: Resultados mostrados en consola.

De esta forma, el reporte incluye tanto la información visual como la numérica, permitiendo una comprensión completa de los resultados generados por el programa.