TP2 Informe

Jose Miguel Alfaro Castillo

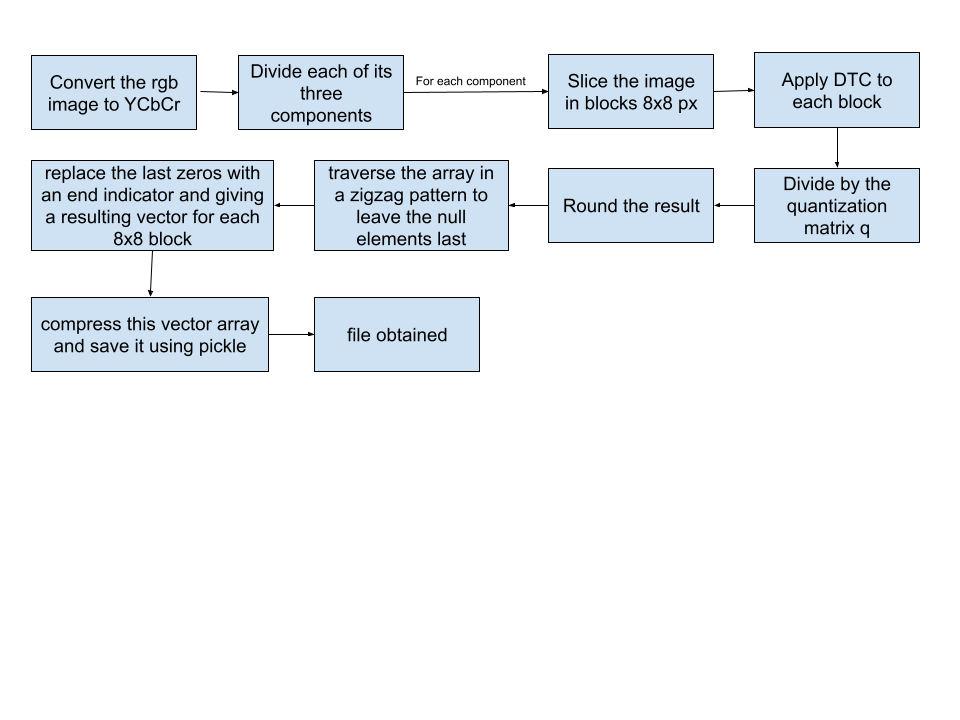
Introducción:

Este TP estuvo dividido en tres partes principales, las cuales fueron, encontrar y seleccionar un algoritmo de compresión de imágenes utilizando DHT, el entrenamiento de diferentes algoritmos de machine learning para clasificar números del 0 – 9 escritos a mano, para posteriormente repetir el proceso despues de preprocesar las imágenes con el HOG (Histogram of Gradient Feature), para posteriormente comparar estos dos resultados y encontrar cual presenta mejor rendimiento. Por ultimo la ultima parte habla sobre el proceso para generar un mosaico de thumbnails, empezando por una large image, se tomaran en en cuenta tres posibles procesos, variando las características tomadas por cada bloque de la imagen y tambien la forma de calcular la distancia entre cada bloque de la imagen y los thumbnails disponibles.

Methodologia:

1. Compresión de imágenes con DHT

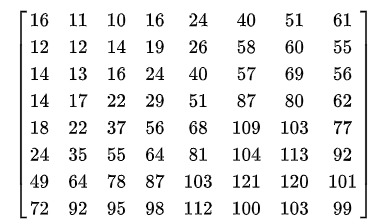
Lo primero fue definir el proceso para comprimir una imagen utilizando el algoritmo DHT, este proceso se ilustra en la siguiente imagen:



El proceso de descompresión es el inverso a este, utilizando la misma q que se utilizo anteriormente.

Para tener un marco de comparación este algoritmo se realizó inicialmente con un q por defecto que se utiliza para la comprimir las imágenes JPEG.

Este q es:



La razón por la que los valores van subiendo a medida que aumentan filas y columnas es porque esta matriz busca generar la mayor cantidad de ceros posibles cuando la matriz de coeficientes DHT sea dividida.  
  
Con la finalidad de encontrar un q que logre tener mejores resultados que la que llamaremos q\_JPEG, se propone un experimento aleatorio donde primero se define una matriz q de la forma:

Q(i,j)=a+b(i+j)

Durante 200 iteraciones, a va a tomar un valor aleatorio entre 3 y 100 , y b un valor aleatorio entre 1 y 40. Con cada uno de estos valores se genera la imagen comprimida llamada q\_a{a}\_b{b}, calculamos su tamaño, la descomprimimos, se obtiene la imagen en RGB al realizar el proceso contrario al mostrado para la compresión y esta imagen es comparable con la imagen comprimida con JPEG, y con la imagen de origen. Con la finalidad de encontrar si es que se encuentran siguiendo este método, algunos valores de a y b que logren obtener una mejor relación de compresión y calidad, tomando como referencia, el tamaño y la diferencia entre la imagen original y la comprimida medida por el indicador PSNR.

1. Clasificación de imágenes:

Para la realización de este punto, de obtuvieron 70000 imágenes etiquetadas de mnist\_784. Con la finalidad de entrenar varios modelos de machine learning y encontrar el más optimo para clasificar esta base de datos.

Lo primero que se hace es normalizar los datos para que estén entre 0,1 y de esta manera que ningún algoritmo de clasificación tenga problemas con su procedimiento. Para entrenar estos datos, fueron utilizados 3 de los métodos vistos en clase, estos fueron KNN, árbol de decisión y SVM y se busca comparar sus rendimientos.

Se entrena y evalua cada uno de estos modelos de machine learning utilizando Kfold partido en 5 y se pondera el resultado de cada uno para compararlos.  
  
Como segunda parte del procesamiento de imágenes, se propone la utilización del HOG (Histogram of Gradient Feature). Inicialmente con la finalidad de comprender como este trabaja se le aplicó a una de las imágenes PPM disponible y se ilustra tanto el histograma de pesos de la imagen como una imagen que ilustra el HOG para cada uno de los bloques.  
Se procede a utilizar un preprocesamiento con HOG a todos los datos provenientes de mnist\_784 y posteriormente se repite el entrenamiento de los modelos utilizando estas características de HOG y no todos los datos de la imagen.  
  
Se compara el rendimiento de cada uno de los modelos posterior a utilizar el HOG para encontrar si hay una mejoría en la precisión de los modelos.

Por ultimo se nos propone la pregunta de que podemos esperar que suceda si se hiciera el proceso con imágenes comprimidas con JPEG QF=75. Lo que es más esperable es que clasifique peor las imágenes al llegar a pixelearlas debido que recorta demasiada información, para estar seguros de esto, se hará la prueba con los datos y se repetirá el proceso de entrenamiento para encontrar la respuesta.

1. Large Mosaic:

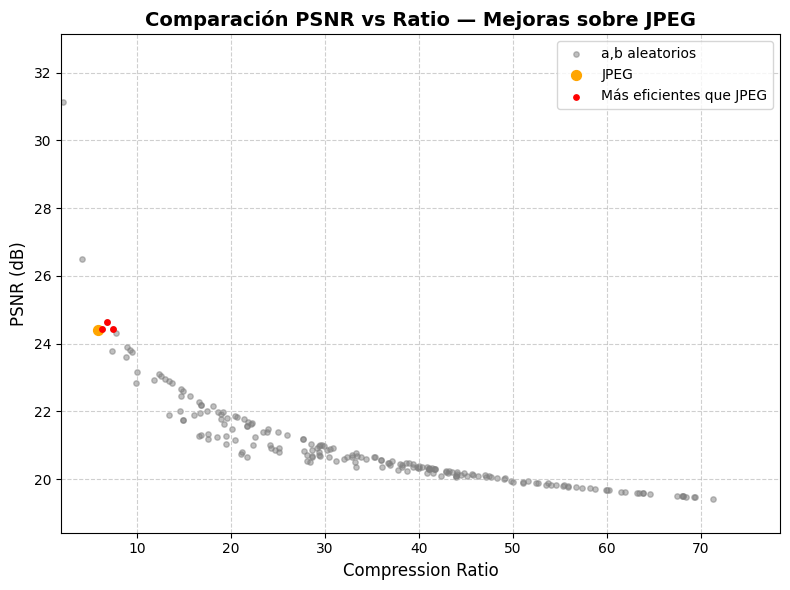
La idea es generar un mosaico de imágenes pequeñas, teniendo de referencia una imagen grande, logrando que se identifiquen las imágenes pequeñas sin perder el marco general de la imagen.

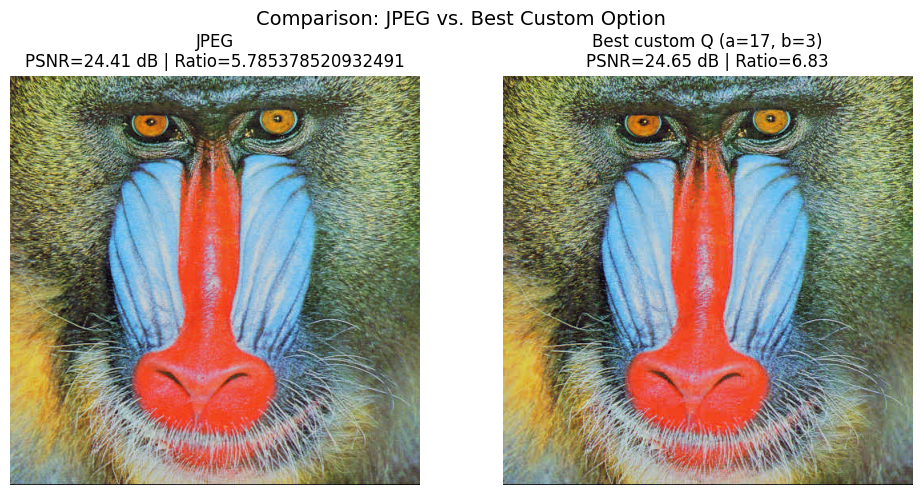
Lo primero que se hará es seleccionar una imagen grande y seleccionar un tamaño para reemplazar la imagen por bloques, teniendo en cuenta la finalidad de lograr que se pueda ver la idea general de la imagen grande sin perder de la vista las más pequeñas, por lo que se seleccionó un tamaño de bloque de 32x32, el tamaño actual de las imágenes pequeñas. Se redimensiona la imagen grande para que puedan reemplazarse todos sus pixeles más fácilmente.

Lo siguiente es seleccionar las características que describirán las imágenes pequeñas, se tomarán tres medidas diferentes, inicialmente la media de rgb, en segunda instancia el histograma HSV y por ultimo un método que pondera estos.  
Lo siguiente es seleccionar que medida de distancia se utilizará para encontrar el thumbnail más cercano a un determinado bloque de la imagen. Se tendrán en cuenta principalmente dos distancias, euclideana y cityblock.

Finalmente se compararán los tres resultados obtenidos para sacar conclusiones de como las características tomadas y las distancias, pueden influir en el desempeño.

Analisis de resultados:  
  
1. Compresión de imágenes con DHT:

Para considerar alguno de los resultados como más eficiente que JPEG, debe lograr conseguir una tasa de compresión más alta que este, mientras que la calidad de la imagen medida con PSNR, sea igual o mejor que JPEG o que la calidad de la imagen sea mejor que la de JPEG con una compresión igual o mayor.  
  
Teniendo en cuenta esta condición se realizaron las 200 pruebas aleatorias obteniendo los siguientes resultados:  


Como se puede ilustrar se encontraron tres imágenes con diferentes valores de a y b que consiguen un rendimiento mejor al de JPEG. Esto es teniendo en cuenta el indicador PSNR que es un indicador matemático, como ayuda visual se encuentra la siguiente imagen comparativa.

Por lo menos a simple vista podemos decir que no se encuentran diferencias significativas, pero la imagen de la derecha logra comprimir la imagen inicial consiguiendo hacerla más ligera y obteniendo un PSNR mejor.

2. Clasificación de imágenes

| **Model** | **Average Accuracy** | **Standard Deviation** |
| --- | --- | --- |
| KNN (k=3) | 0.971217 | 0.001026 |
| Decision Tree (Gini) | 0.870100 | 0.001932 |
| SVM (RBF) | 0.983133 | 0.000602 |

Los tres modelos probados permitieron clasificar directamente las imágenes MNIST sin extraer características.

El SVM con núcleo RBF obtuvo la mejor precisión media (97 %), seguido del KNN (96 %), mientras que el árbol de decisión (87 %) dio resultados inferiores, pero más rápidos de entrenar.

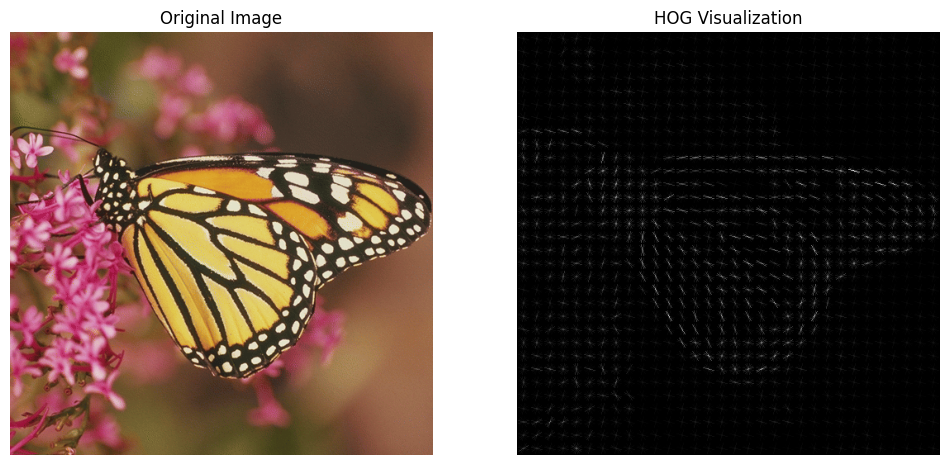
Posteriormente se habría de utilizar el HOG, este se ve de la siguiente manera para una imagen PPM  


Tabla Resultados modelo entrenamiento con HOG

| **Model** | **Average Accuracy** | **Standard Deviation** |
| --- | --- | --- |
| KNN (k=3) | 0.973717 | 0.001026 |
| Decision Tree (Gini) | 0.868717 | 0.001932 |
| SVM (RBF) | 0.989383 | 0.000602 |

Tabla de resultados base sin preprocesamiento vs preprocesamiento con HOG

| **Model** | **Average Accuracy** | **Average Accuracy** | **Standard Deviation** |
| --- | --- | --- | --- |
| KNN (k=3) | 0.971217 | 0.973717 | 0.001026 |
| Decision Tree (Gini) | 0.870100 | 0.868717 | 0.001932 |
| SVM (RBF) | 0.983133 | 0.989383 | 0.000602 |

1. Desempeño Superior del SVM (RBF)

El modelo SVM (Support Vector Machine) con kernel RBF muestra consistentemente la mejor precisión (0.983133 sin HOG y 0.989383 con HOG) y la desviación estándar más baja (0.000602).

Kernel RBF (Radial Basis Function): Al usar el kernel RBF, el SVM es capaz de mapear los datos a un espacio de mayor dimensión donde las clases pueden ser separables linealmente (o casi). Esto le permite manejar las relaciones no lineales y complejas inherentes a las características de las imágenes.

Baja Desviación Estándar: La desviación estándar muy baja indica que el rendimiento del SVM es extremadamente estable y consistente en diferentes particiones de los datos de entrenamiento (p. ej., en la validación cruzada). Esto confirma que el modelo no está sobreajustado y tiene una gran capacidad de generalización.

2. El Impacto de HOG

El Histogram of Oriented Gradients (HOG) es un descriptor de características diseñado para capturar la forma y estructura de un objeto codificando la distribución de los gradientes de intensidad (es decir, los bordes y contornos) en una imagen.

Aumento de la Precisión: HOG convierte los píxeles brutos de la imagen en un vector de características más informativo y compacto.

Mejora Específica en SVM y KNN: Para el SVM, proporcionarle características bien definidas y menos "ruidosas" como las de HOG le permite encontrar un hiperplano de separación aún mejor

3. Mosaico de imágenes:  
La imagen Larga inicial es la siguiente:  

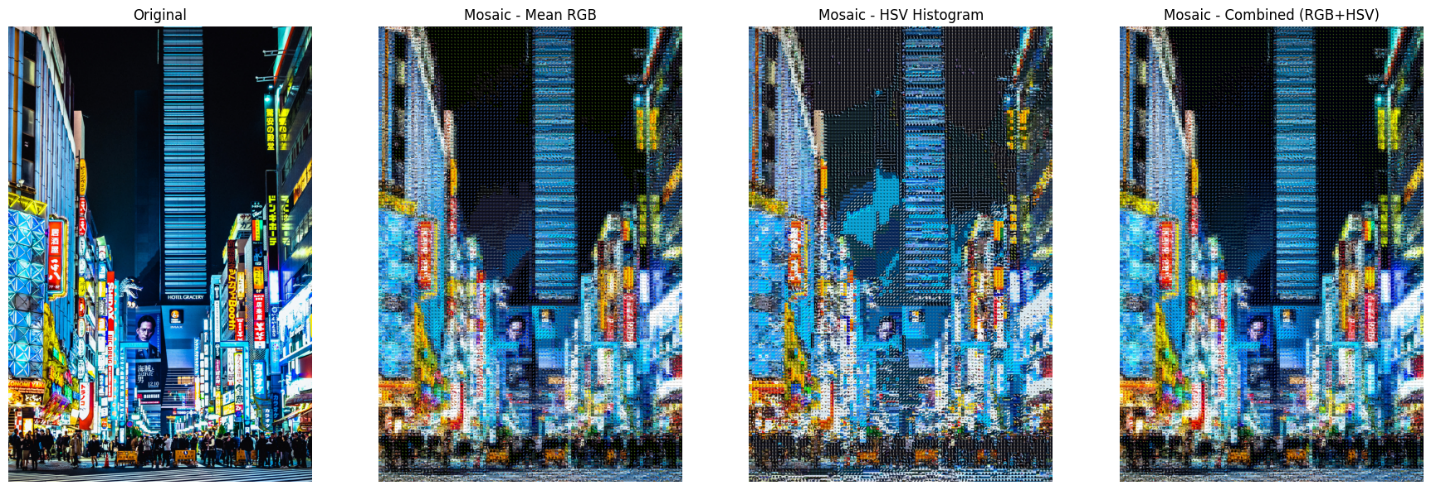

Tomando como base esta imagen y utilizando los thumbnails provenientes de cifar10, se utilizaron tres métodos para convertirlo en un mosaico, los resultados de eficiencia de cada uno son los siguientes:

|  |  |  |  |
| --- | --- | --- | --- |
| Method | PSNR | SSIM | ΔE (avg) |
| Mosaic - Mean RGB | 11.531 | 0.117 | 19.010 |
| Mosaic - HSV Histogram | 10.022 | 0.145 | 22.124 |
| Mosaic - Combined (RGB+HSV+Adjust) | 11.622 | 0.131 | 18.411 |

Best PSNR: Mosaic - Combined (RGB+HSV+Adjust)

Best SSIM: Mosaic - HSV Histogram

Best Color Match (lowest ΔE): Mosaic - Combined (RGB+HSV+Adjust)



A simple vista parece ser la opción obtenida por HSV un poco menos precisa, sin embargo, según el indicador de SSIM se trata de la mejor opción,

¿Por qué pasa esto?

El método HSV puede haber elegido thumbnails con **patrones de luminancia o textura** más parecidos al bloque original, aunque con **colores incorrectos**.

Visualmente, nuestro cerebro penaliza el error de color mucho más que SSIM, que sólo ve que las “formas” o “relieves” son parecidos.

Por eso parece peor a ojo, pero matemáticamente “más estructuralmente similar”.

Finalmente Según el indicador ΔE El método combinado logró la mayor similitud con la imagen base este busca compararla fidelidad cromática y de esta manera es un índice que puede estár correctamente ligado a que se vea más parecido.

Conclusiones:

1. **Compresión de imágenes con DHT**

* El método de compresión implementado a partir de la DHT demostró ser eficaz al conseguir comprimir la imagen sin perdidas de calidad visibles
* Los experimentos aleatorios con distintos parámetros de y permitieron identificar configuraciones que superan al estándar JPEG en términos de PSNR, manteniendo una relación de compresión similar o incluso mejor. Es importante resaltar que esta fue una de las maneras posibles para encontrar una matriz Q, para próximos experimentos recomendaría probar con una prueba aleatoria para muchas más gamas de valores que los que permite una ecuación de la forma .
* Esto muestra que la estructura clásica de cuantificación tipo JPEG no es necesariamente óptima para todas las imágenes, y que ajustes adaptativos de los coeficientes pueden mejorar la fidelidad visual sin aumentar el tamaño del archivo.

1. **Clasificación de imágenes (MNIST)**

* Entre los modelos probados (KNN, Árbol de Decisión y SVM), el SVM con kernel RBF ofreció el mejor desempeño, alcanzando una precisión media de 98.9 % y la menor desviación estándar.
* La aplicación del descriptor HOG permitió mejorar ligeramente los resultados, evidenciando que la extracción de características estructurales aporta información más relevante que los valores de píxeles crudos.
* En general se puede afirmar que HOG tiende a mejorar el rendimiento a la hora de analizar imágenes al brindar características estructurales que aporta una información importante al modelo.

1. **Mosaico de imágenes**

* De los tres enfoques probados, el método Combinado (RGB + HSV + Adjust) ofreció el mejor compromiso global entre fidelidad cromática (menor ΔE), estructura (SSIM) y relación de color (PSNR).
* En este caso HIV consiguió un SSIM alto, pero visualmente no logra tan buena similitud como los otros dos métodos, puede deberse a que se enfoca demasiado en la estructura local.