

Práctica 3

Visión por Computador:

Técnicas de indexación y recuperación de imágenes
usando modelos de bolsa de palabras.

Óscar David López Arcos, 75571640-B
odlarcos@correo.ugr.es
Grupo: Miércoles 11:30-13:30

Índice

Ejercicio 1	3
Ejercicio 2	4
Ejercicio 3	6

Ejercicio 1

Emparejamiento de descriptores

En este primer ejercicio he tenido que introducir los puntos a mano, ya que las funciones dadas para la extracción no funcionaban correctamente en mi sistema operativo.

Para cada par de imágenes seleccionadas, he construido una máscara del tamaño de la imagen original con valores distintos de 0 para aquellos píxeles que se encuentren dentro del polígono representado por los puntos señalados (**fillConvexPoly**).

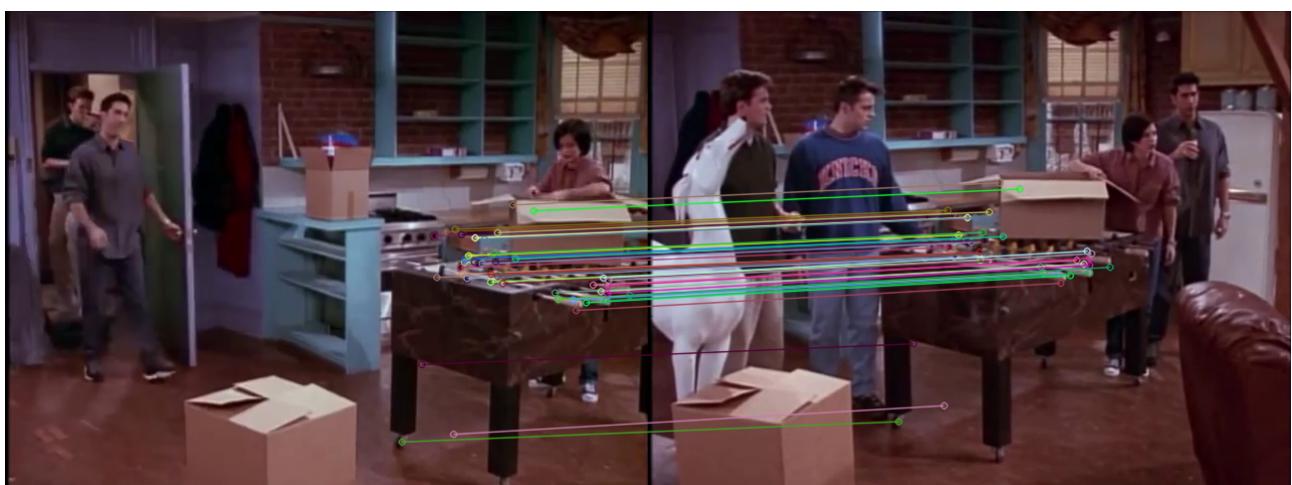
Después, de forma similar a la práctica 2, calculamos las correspondencias entre ambas imágenes pero solo para la zona delimitada por la máscara. He utilizado el algoritmo Lowe-Average-2NN.

Los resultados obtenidos han sido:

Para los objetos de la mesa, en las imágenes 407 y 408:



Para el futbolín, en las imágenes 128 y 130:



Para el logotipo de Central Perk, en las imágenes 429 y 428:



Como conclusión, podemos afirmar que la recuperación a partir de descriptores solamente es útil cuando los objetos en ambas imágenes aparecen en posición muy parecida. Los descriptores SIFT son bastante sensibles a transformaciones proyectivas ya que dependen del gradiente de los píxeles. Es por eso que en la última imagen, a pesar de ser el mismo logotipo, le ha sido imposible encontrar ningún punto en común ya que la perspectiva y condiciones lumínicas son totalmente diferentes.

Esta técnica solo debería utilizarse en situaciones muy controladas, donde tengamos la certeza de la similitud entre las imágenes.

Ejercicio 2

Recuperación de imágenes

Este ejercicio nos pide recuperar imágenes similares a una imagen-pregunta, basándonos en el criterio de bolsa de palabras.

Para este proceso, he definido la función **imagenes_similares** que realizará las siguientes tareas:

- Calcular el histograma para la imagen_pregunta (**calcular_histograma2**).
- Calcular los histogramas del resto de imágenes.
- Crear la tabla de índices invertidos* (**inverted_index**)
- Calculamos la similitud entre los histogramas de la imagen pregunta y el resto. (**similitud**)
- Escoger las 5 imágenes con mayor similitud.

Calcular Histograma

El cálculo del histograma de las imágenes lo afronté de dos formas distintas. En un primer momento (**calcular_histograma**), intenté calcular la distancia euclidea para cada descriptor de la imagen con todos las palabras del diccionario para quedarme con la más cercana.

Este proceso era muy lento, por lo que finalmente decidí aplicar el algoritmo de Fuerza Bruta de la práctica 2 para calcular las correspondencias entre los descriptores y las palabras (**calcular_histograma2**). Al hacerlo sin validación cruzada, asignará a cada punto la palabra más cercana, justo como necesitamos. Por último construiremos el histograma en función de las veces que aparezca cada palabra en la imagen

Similitud

Aplicando la fórmula vista en clase entre ambos histogramas, obtenemos la similitud (**similitud**).

Índice invertido

Para la construcción del índice invertido, para cada palabra iteramos sobre todos los histogramas y, si esa palabra aparece en el histograma de una imagen (valor distinto de 0), añadimos dicha imagen a la lista en el índice. Este cálculo ralentizaba mucho el proceso de obtención de imágenes similares, así que como no es necesario para el ejercicio, he dejado comentadas ambas líneas.

Veamos a continuación un caso donde funciona bien este modelo:



Imagen Pregunta



Imágenes similares

Ahora, veamos un caso donde realmente falla:



Imagen Pregunta



Imágenes similares

Como se puede observar a la vista de los resultados, cuando la escena es lo suficientemente distintiva del resto, la recuperación de imágenes de la misma escena siguiendo el modelo de bolsa de palabras funciona adecuadamente, ya que las palabras en estas escenas serán muy particulares y parecidas entre sí.

Sin embargo, para una imagen no tan distintiva, las palabras que aparecen pueden estar presentes en muchas otras imágenes ya que dependen de descriptores obtenidos por el gradiente. Al tener todas un tono parecido, es fácil crear confusión y que todas las escenas se parezcan demasiado. Aún así es posible que todas estén relacionadas por palabras referentes a los muebles/sofás, que es lo que aparece en las imágenes.

Ejercicio 3

Visualización del vocabulario

Cargamos los descriptores y patches visuales extraídos de las imágenes. Al haberse construido el vocabulario a partir de estos descriptores, los más similares a cada palabra deberían tener patches de apariencias parecidas.

Cargamos el diccionario y escogemos dos palabras para las cuales calcularemos la distancia euclídea a cada descriptor. Nos quedamos con los más cercanos a cada una (**best_patches**).

El resultado ha sido:



Palabra 728



Palabra 321

Se puede observar que, aunque algunos sí presentan características comunes (como el segundo y quinto Patch de la palabra 728), en general no es la cercanía visual esperada.

Esto viene a resaltar la complejidad de este proceso y lo difícil que es generar un diccionario efectivo para los problemas que hemos abordado.