

Práctica 3 - Visión por computador

Fecha de entrega: 22 de diciembre

Información de la entrega de prácticas

- La entrega se realiza a través del tablón docente de la plataforma DECSAI.
- Valoración total: 11 puntos (+ 4 puntos de bonus)

Informe a presentar

Tanto para este trabajo como para los demás proyectos debe presentar un informe escrito con sus valoraciones y decisiones adoptadas en cada uno de los apartados de la implementación. También deberá incluirse una valoración sobre la calidad de los resultados encontrados.

Normas de la entrega de prácticas

El incumplimiento de estas normas significa la pérdida directa de un punto cada vez que se detecte un incumplimiento.

1. El código se debe estructurar en funciones, una por cada apartado de la práctica.
2. El código debe estar obligatoriamente comentado explicando lo que realizan los distintos apartados y/o bloques.
3. Todos los ficheros juntos se podrán dentro de un fichero zip, cuyo nombre debe ser `Apellido1_P[1-3].zip`.
4. Los path que se usen en la lectura de imágenes o cualquier fichero de entrada debe ser siempre `imagenes/nombre_fichero`.
5. Todos los resultados serán mostrados por pantalla. No escribir nada en el disco.
6. La práctica debe poder ejecutarse de principio a fin sin necesidad de ninguna selección de opciones. Para ello, se deben fijar los parámetros por defecto que se consideren óptimos.
7. Solo poner puntos de parada para mostrar imágenes o datos por consola.

Ayudas

- El fichero `archivosP3.zip` (disponible a través del enlace) contiene los archivos:
 - `imagenesIR.rar` con las imágenes necesarias para la práctica.
 - `descriptorsAndpatches2000.pkl` con los descriptores y parches asociados.
 - `kmeanscenters2000.pkl` con el vocabulario dado.
 - El archivo `auxFunc.py` con código auxiliar.
- Los ficheros `descriptorsAndpatches2000.pkl` y `kmeanscenters2000.pkl` se pueden leer usando `loadAux()` y `loadDictionary()` del fichero `auxp3.py`.
- Las funciones `extractRegion()` y `clickAnddraw()` disponibles en el archivo `auxFunc.py` permiten fijar una región de forma interactiva, dando como salida las coordenadas de los vértices del polígono.
- Usar la función `cv2.kmeans()` para la construcción del vocabulario. En clase de Prácticas se explicará su uso.
- Para una correcta visualización de los parches asociados a cada palabra visual, estos deberán ser re-escalados a un tamaño prefijado. Usar la función `cv2.resize()` (p.e 16x16 o 24x24).
- En Windows, se deben presentar las imágenes de parches usando `cv2.imshow()` y no a través de `pyplot` (las imágenes son demasiado pequeñas).
- Se aconseja normalizar los vectores descriptores para que tengan norma euclídea 1.

Trabajo de implementación

Este trabajo está dirigido a implementar técnicas de indexación y recuperación de imágenes usando modelos de bolsa de palabras. Para ello haremos uso de las características SIFT extraídas de las imágenes, calcularemos un diccionario en el que representar dichas características y haremos uso del modelo de bolsa de palabras para caracterizar y comparar imágenes o regiones entre sí.

1. Emparejamiento de descriptores [4 puntos]
 - Mirar las imágenes en `imagenesIR.rar` y elegir parejas de imágenes que tengan partes de escena comunes. Haciendo uso de una máscara binaria o de las funciones `extractRegion()` y `clickAndDraw()`, seleccionar una región en la primera imagen que esté presente en la segunda imagen. Para ello solo hay que fijar los vértices de un polígono que contenga a la región.

- Extraiga los puntos SIFT contenidos en la región seleccionada de la primera imagen y calcule las correspondencias con todos los puntos SIFT de la segunda imagen (ayuda: use el concepto de máscara con el parámetro `mask`).
- Pinte las correspondencias encontrados sobre las imágenes.
- Jugar con distintas parejas de imágenes, valorar las correspondencias correctas obtenidas y extraer conclusiones respecto a la utilidad de esta aproximación de recuperación de regiones/objetos de interés a partir de descriptores de una región.

2. Recuperación de imágenes [4 puntos]

- Implementar un modelo de índice invertido + bolsa de palabras para las imágenes dadas en `imagenesIR.rar` usando el vocabulario dado en `kmeanscenters2000.pkl`.
- Verificar que el modelo construido para cada imagen permite recuperar imágenes de la misma escena cuando la comparamos al resto de imágenes de la base de datos.
- Elegir dos imágenes-pregunta en las se ponga de manifiesto que el modelo usado es realmente muy efectivo para extraer sus semejantes y elegir otra imagen-pregunta en la que se muestre que el modelo puede realmente fallar. Para ello muestre las cinco imágenes más semejantes de cada una de las imágenes-pregunta seleccionadas usando como medida de distancia el producto escalar normalizado de sus vectores de bolsa de palabras.
- Explicar qué conclusiones obtiene de este experimento.

3. Visualización del vocabulario [3 puntos]

- Usando las imágenes dadas en `imagenesIR.rar` se han extraído 600 regiones de cada imagen de forma directa y se han re-escalado en parches de 24x24 píxeles. A partir de ellas se ha construido un vocabulario de 5.000 palabras usando k-means. Los ficheros con los datos son `descriptorsAndpatches2000.pkl` (descriptores de las regiones y los parches extraídos) y `kmeanscenters2000.pkl` (vocabulario extraído).
- Elegir al menos dos palabras visuales diferentes y visualizar las regiones imagen de los 10 parches más cercanos de cada palabra visual, de forma que se muestre el contenido visual que codifican (mejor en niveles de gris).
- Explicar si lo que se ha obtenido es realmente lo esperado en términos de cercanía visual de los parches.

Bonus

Sólo se tendrán en cuenta si se ha obtenido más del 70% de los puntos obligatorios.

1. Recuperación de regiones [2 puntos]: Usando la misma función del Ejercicio 1, seleccione una región rectangular que le parezca relevante de una imagen del conjunto **imagenesIR.rar** y calcule su modelo de bolsa de palabras ponderando las palabras con los pesos calculados por el criterio tf-idf. Use el diccionario dado o uno propio calculado. Además:
 - (a) Use la imagen seleccionada para extraer las 20 imágenes más cercanas a ella en imagenesIR. (usar el Ejercicio 2). Verifique que en un porcentaje razonable de ellas aparece la región seleccionada aunque posiblemente con otros contextos.
 - (b) Recorrer las 20 imágenes usando un modelo de ventana deslizante y calcule en cada posición de la ventana un modelo de bolsa de palabras. Mida la distancia de los modelos de bolsa de palabras (ventana y región seleccionada) y guarde las coordenadas de la región y su distancia asociada. Muestre en orden las 20 regiones más cercanas a la región seleccionada. ¿Cómo influye el tamaño de la ventana deslizante en los resultados de la detección? ¿Qué conclusiones extrae de sus resultados?
2. Creación de un vocabulario [2 puntos]: Calcular desde todas las **imagenesIR.rar** los ficheros dados en el Ejercicio 2 usando los mismos parámetros. Aplicar con el nuevo diccionario lo pedido con el Ejercicio 3.