

Trabajos de Visión por Computador

TRABAJO-3

Indexación y Recuperación de Imágenes

FECHA DE ENTREGA: 22 de diciembre

Valoración total: 17 puntos

Informe a presentar

Para este trabajo como para los demás proyectos es obligatorio presentar un informe escrito comentando los detalles del desarrollo de cada ejercicio junto 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. (hacer en pdf o texto plano)

Normas de la entrega de Prácticas: EL INCUMPLIMIENTO DE ESTAS NORMAS SIGNIFICA PERDIDA DIRECTA DE 1 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 numéricos serán mostrados por pantalla. No escribir nada en el disco.
6. La práctica deberá poder ser ejecutada de principio a fin sin necesidad de ninguna selección de opciones. Para ellos fijar los parámetros por defecto que se consideren óptimos.
7. Solo poner puntos de parada para mostrar imágenes o datos por consola

Cuestionario de Teoría: (estará disponible en la web más adelante)

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 [3 puntos]: Leer parejas de imágenes de imagenesIR.rar que tengan partes de escena comunes (p.e. (128,130), (143,145), (229,232), etc). 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 (ver explicación más abajo). 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 del parámetro " mask"). Pinte las correspondencias encontrados sobre las imágenes. Jugar con distintas parejas de imágenes y decir que conclusiones se extraen de los resultados obtenidos con respecto a la utilidad de esta aproximación en la recuperación de imágenes a través de descriptores.

2.- Visualización del vocabulario [3 puntos] : Usando las imágenes dadas en imagenesIR.rar se han extraído 600 regiones de cada imagen y se ha construido un vocabulario de 5.000 palabras usando k-means. Se han extraído de forma directa las regiones imágenes asociadas y se han re-escalado a 24x24 píxeles. Los ficheros con los datos son descriptors.pkl, vocabulary.pkl y patches.pkl. Leer los ficheros usando loadAux() y loadDictionary(). Elegir al menos dos palabras visuales diferentes y visualizar las regiones imagen de los 20 parches más cercanos de cada palabra visual, de forma que se muestre el contenido visual que codifican. Explicar lo que se ha obtenido.

3.- Recuperación de imágenes [3 puntos]: Implementar un modelo de índice invertido + bolsa de palabras para las imágenes dadas en imágenesIR.rar usando el vocabulario calculado en el punto anterior. 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 los resultados.

BONUS: (solo 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 apartado.1 seleccione una región rectangular que le parezca relevante de una imagen del conjunto imagenesIR.zip y calcule su modelo de bolsa de palabras ponderando las palabras con los pesos calculados por el criterio tf-idf.

- a) Use esa imagen para extraer las 20 imágenes más cercanas a ella en la base de datos usando el punto anterior y verifique que en un porcentaje razonable de ellas aparece la región seleccionada aunque en otros contextos. Use un esquema de ventana deslizante sobre las 20 imágenes seleccionadas usando el mismo tamaño de la región seleccionada.
- b) Recorra 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. Que conclusiones extrae de sus resultados.

2.-Consistencia espacial [2 puntos]. Implementar la verificación espacial afín a través del método RANSAC al problema del bonus anterior. Probar con ejemplos la mejora que supone frente al no uso de verificación espacial. (Leer el paper de Sivic

and Zisserman <http://www.robots.ox.ac.uk/~vgg/publications/papers/sivic09a.pdf> para entender bien la técnica)

3.- Creación de un vocabulario [2 punto]: Calcular desde imágenesIR.rar los ficheros dados en el punto-2

Forma de entrega:. Subir el zip al Tablón docente de CCIA.

Ayudas: Están disponibles algunas funciones que pueden ser usadas en el desarrollo de la práctica, aunque no son necesarias

- 1.- Las funciones `extractRegion()`+ `clickAnddraw()` permiten fijar una región de forma interactiva, dando como salida las coordenadas de los vértices del polígono
- 2.- Usar la función `cv2.kmeans()` para la construcción del vocabulario. En clase de Prácticas se explicará su uso.
- 3.- 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)
- 4.- Los que usen Windows, deben presentar las imágenes de parches usando `cv2.imshow()` y no a través de `pyplot` (imágenes demasiado pequeñas)
- 5.- Se aconseja normalizar los vectores descriptores a tengan norma cuadrática igual a 1.