

Práctica 3

Visión por Computador

Ignacio Aguilera Martos

22 de Diciembre de 2018

Índice

1. Ejercicio 1	2
2. Ejercicio 2	6
3. Ejercicio 3	6

1. Ejercicio 1

Emparejamiento de descriptores.

- 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 sólo 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 encontradas 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.

Solución

Las parejas que he escogido han sido dos para exemplificar el buen comportamiento cuando las imágenes son similares entre sí y otra en la que el reconocimiento no es tan bueno.

Las parejas escogidas son:



Figura 1: Imágenes 1 y 4



Figura 2: Imágenes 23 y 24

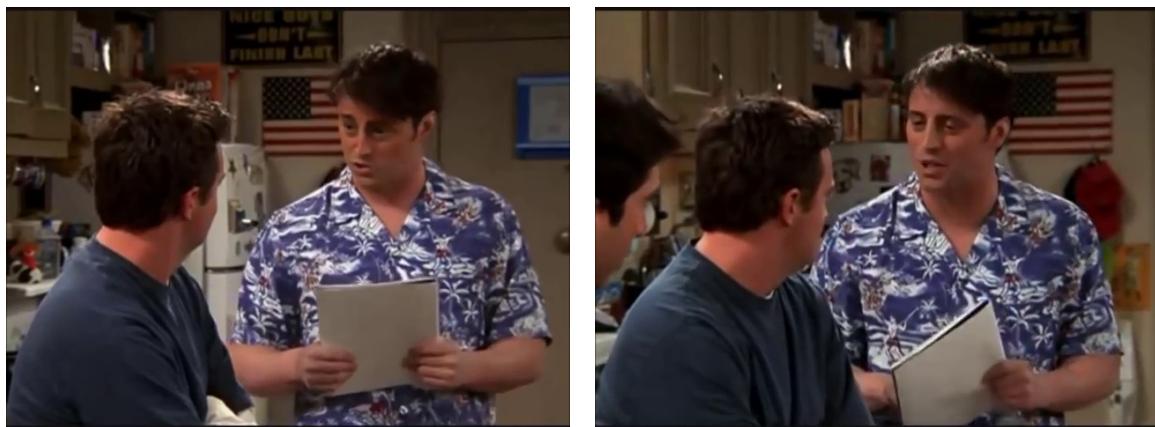


Figura 3: Imágenes 91 y 92

La forma de proceder es, mostrar la imagen para que se pueda seleccionar la región de la misma que se desee. Esta región es diferenciada del resto mediante una máscara. La construcción de dicha máscara se hace tomando los puntos que determinan el polígono de la región y, mediante la función `fillConvexPoly` se rellena esta área a blanco, esto es `(255, 255, 255)`. Tras esto sólo tenemos que crear una matriz de ceros del mismo tamaño que la imagen y, en dichos puntos poner el valor 1. Con esto tendríamos una máscara que podemos aplicar a la función `detectAndCompute` para hallar los keypoints y descriptores sólo de la región que engloba el polígono.

Hay que tener en cuenta que la detección puede ser buena si realmente hay un parecido importante entre la región que hemos seleccionado y alguna región de la imagen con la que queremos emparejar. Por ejemplo veamos los dos primeros ejemplos y analicémoslos.

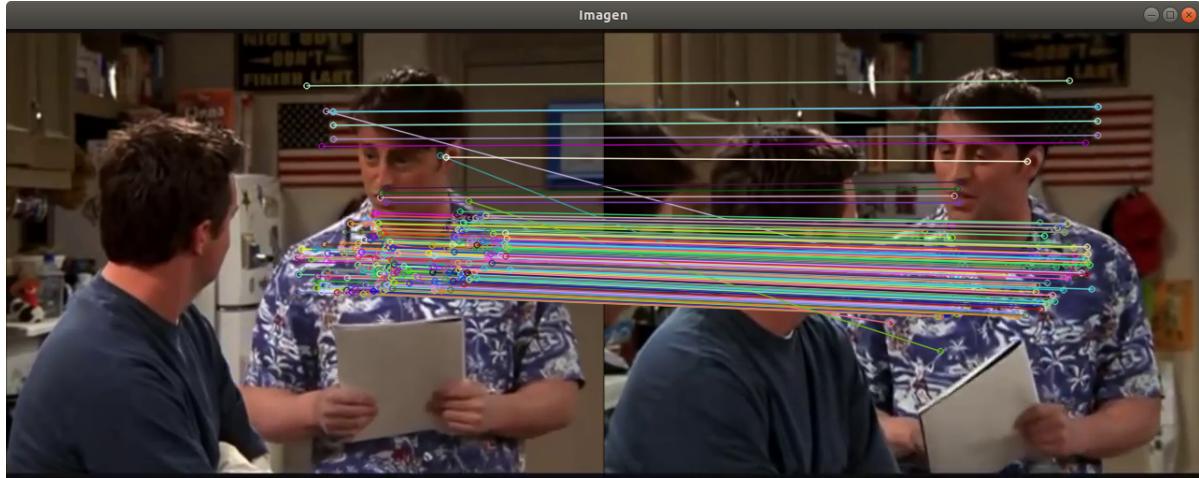


Figura 4: Emparejamiento con las imágenes 91 y 92



Figura 5: Emparejamiento con las imágenes 23 y 24

Como podemos ver claramente en estas dos imágenes el emparejamiento es muy bueno puesto que los parecidos entre las dos es muy alto. Es prácticamente la misma cara con la misma escena y el mismo fondo, por lo que no sólo se empareja la cara o la figura de la persona, si no, también el fondo como en el segundo de los casos con el cuadro que se encuentra tras la persona.

Cabe decir que esta detección puede variar enormemente como en el siguiente caso paradigmático.

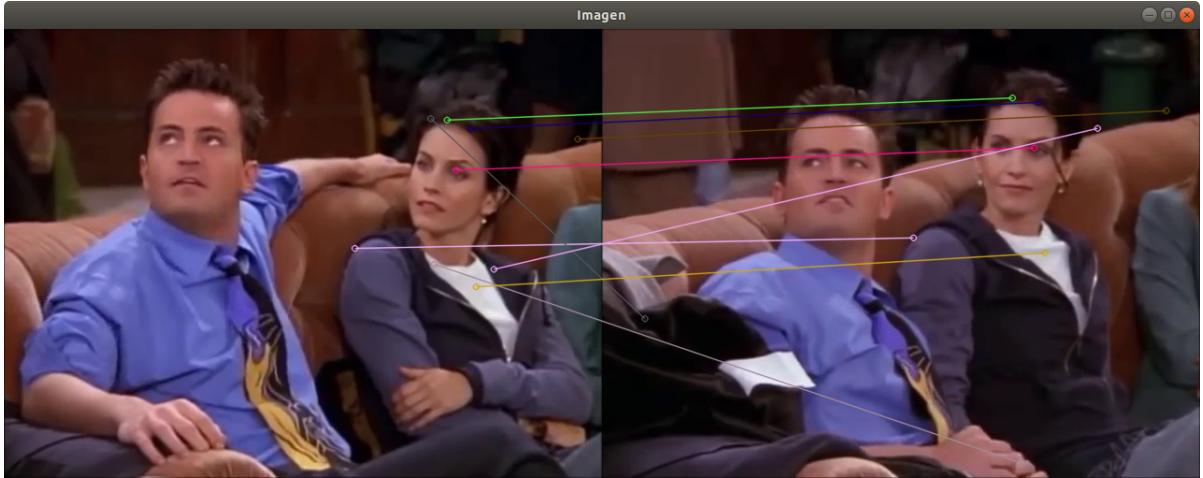


Figura 6: Emparejamiento con las imágenes 1 y 4

En este caso podemos ver que la detección es mucho peor pero analicemos por qué. Realmente si observamos cada uno de los keypoints como se relacionan con la imagen de la derecha vemos que el emparejamiento de los mismos es razonable salvo dos de ellos, pero este emparejamiento es muy pobre porque no hemos obtenido suficientes puntos de interés en la región como para obtener un emparejamiento sólido como en los dos primeros casos.

Ahora bien, podemos tomar también dos imágenes de una escena que comparten una figura como por ejemplo la mujer que se enseña en las imágenes siguientes:



Figura 7: Imágenes 142 y 143

Veamos que la detección en este caso no es satisfactoria cuando tomamos el polígono sobre la figura de la mujer.

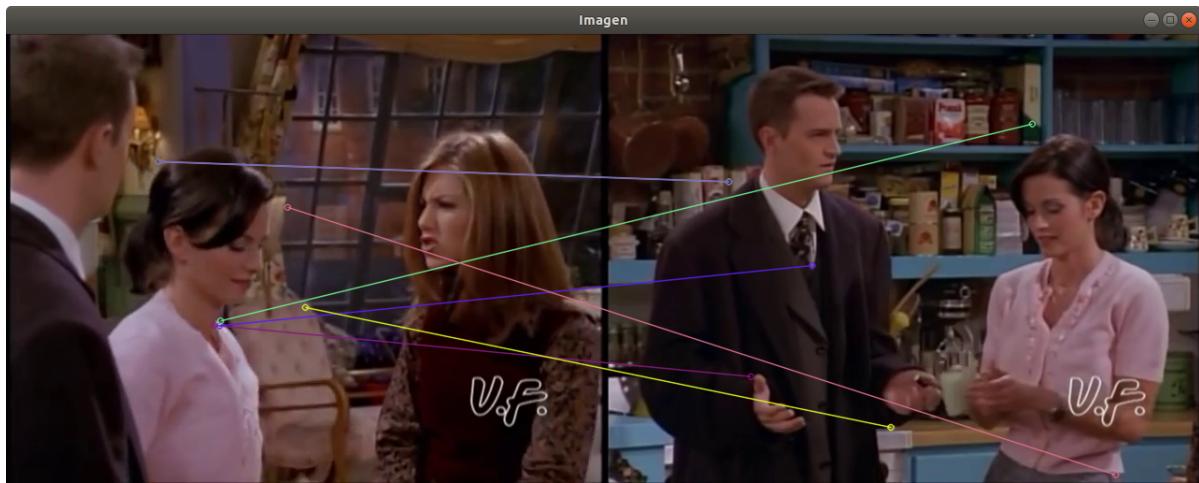


Figura 8: Emparejamiento con las imágenes 142 y 143

Como podemos observar nosotros estamos viendo que la mujer que aparece en la imagen de la izquierda es la misma que en la parte de la derecha pero la detección asigna los puntos de forma completamente errónea con respecto a la figura de la misma. Este fallo viene sobre todo dado porque el fondo de las imágenes es completamente distinto y por tanto, los puntos de interés tomados (que hacen el contraste en este caso de la mujer con el fondo) se ven completamente irrelevantes con respecto a la correlación que hace de los mismos.

Por lo tanto podemos concluir que el uso de esta aproximación a la recuperación de regiones de interés a partir de descriptores de una región es en general útil pero conociendo las limitaciones, tanto de la obtención de un número de puntos de interés demasiado bajo como de el emparejamiento erróneo que pueda realizar la misma con otra imagen en la que sí comparten una región pero el fondo o la iluminación han cambiado y no se perciben las regiones como similares.

2. Ejercicio 2

Recuperación de Imágenes

- 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 que 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.

Solución

3. Ejercicio 3

Visualización del vocabulario

- 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 5000 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 de las 10 imágenes más cercanas 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.

Solución