Visión por Computador

Cuestionario de teoría - 3

Curso 2017/2018

Francisco Javier Caracuel Beltrán

caracuel@correo.ugr.es

4º - Grado en Ingeniería Informática – CCIA – ETSIIT – UGR

Índice

1. ¿Cuáles son las propiedades esenciales que permiten que los modelos de recuperación de instancias de objetos de una gran base de datos a partir de descriptores sean útiles? Justificar la respuesta.

Las propiedades esenciales que permiten recuperar instancias de objetos de una base de datos a partir de descriptores vienen dadas gracias las características de los puntos SIFT.

Los puntos SIFT se mantienen invariantes ante rotaciones y/o escalados de las imágenes, permitiendo manejar cambios en el punto de vista, cambios en la iluminación y siendo una técnica fácil de implementar y eficiente.

2. Justifique el uso del modelo de bolsa de palabras en el proceso de detección y reconocimiento de instancias de objetos. ¿Qué ganamos? ¿Qué perdemos? Justificar la respuesta.

El modelo de bolsa de palabras es un sistema que permite, sin saber de manera certera las palabras óptimas que deben formar la bolsa de palabras, obtener una muy buena aproximación de la imagen sobre la que se realiza la consulta. Es una técnica que funciona bien con un resumen compacto del contenido de la imagen, es flexible a la geometría, deformaciones y al punto de vista. Se gana poder reconocer instancias de manera sencilla y rápida. Se pierde obtener las instancias óptimas de una imagen y su funcionalidad cuando, por ejemplo, la bolsa cubre toda la imagen y engloba el primer plano con el fondo.

3. Describa la diferencia esencial entre los problemas de reconocimiento de instancias y reconocimiento de categorías. ¿Qué deformaciones se presentan en uno y otro? Justificar la respuesta.

La diferencia esencial entre los problemas de reconocimiento de instancias y de reconocimiento de categorías es que las instancias son invariantes entre sí. Una instancia será siempre ella misma, mientras que las categorías representan para los humanos un objeto que cumple una misma finalidad, pero que puede tener multitud de formas que, para los algoritmos, no tienen relación ninguna.

Las deformaciones que podrían tener las instancias son rotaciones o escalados, por lo que utilizar los puntos SIFT sería una buena opción para su detección.

Las deformaciones que se presentan en las categorías pueden ser de cualquier tipo, ya que, por ejemplo, una lámpara puede ser completamente plana o puede ser colgante y no guardan ninguna relación geométrica entre sí.

4. ¿Es posible usar el modelo de bolsa de palabras para el reconocimiento de categorías de objetos? Justificar la contestación.

El modelo de bolsas de palabras se utiliza para encontrar patrones, objetos o elementos de una imagen dada en una base de datos donde existe multitud de imágenes (parecidas o no). Con el modelo de bolsa de palabras, si se recibe una imagen que contiene un paisaje, se podrían encontrar múltiples paisajes que se parecieran a éste, pero, como se ha expuesto en el ejercicio anterior, dos objetos de una misma categoría podrían no guardar relación entre sí, por lo que no serviría como una técnica válida.

5. Suponga que desea detectar, en una imagen, una instancia de un objeto a partir de una foto del mismo, tomada desde el mismo punto de vista del que aparece en la imagen y en un entorno de iluminación similar. Analice la situación en el contexto de las técnicas de reconocimiento de objetos e identifique qué algoritmo concreto aplicaría que fuese útil para cualquier objeto. Argumente por qué funcionaría y especifique los detalles necesarios que permitan entender su funcionamiento.

Cuando se tiene un objeto invariante en dos escenas diferentes, se pueden calcular los descriptores de ambas imágenes. Una vez que se tienen los puntos más relevantes, se utiliza cualquier técnica válida que permita hacer *match* entre los descriptores de ambas imágenes como la vista en clase *Fuerza Bruta + Validación Cruzada*. Una vez realizada esa operación, ya se tendrían disponibles los puntos de la segunda imagen que delimitan el objeto.

Este mismo proceso se ha realizado en el trabajo3, ejercicio 1. La diferencia entre el caso expuesto en este ejercicio y el del trabajo 3 es que no es necesario enviar una máscara que delimite la zona de donde se quieren extraer los descriptores del objeto a buscar, sino que la imagen en sí es el objeto.

Esta técnica funcionaría, aunque el objeto en la escena estuviera rotado o escalado, gracias a las propiedades de los puntos SIFT. Esta técnica permite encontrar el objeto gracias a que no varía su forma geométrica.

6. Suponga de nuevo el problema del ejercicio anterior pero la foto que le dan está tomada con un punto de vista del objeto distinto respecto del objeto en la imagen. Analice qué repercusiones introduce esta modificación en su solución anterior y qué cambios debería de hacer para volver a tener un nuevo algoritmo exitoso. Justificar la respuesta.

Con la técnica expuesta en el ejercicio anterior no sería necesario modificar nada, ya que el objeto como tal, sigue siendo el mismo, aunque su enfoque varíe.

7. Suponga que una empresa de Granada le pide implementar un modelo de recuperación de información de edificios históricos de la ciudad a partir de fotos de los mismos. Explique de forma breve y clara qué enfoque le daría al problema. ¿Qué solución les propondría? Y, ¿cómo puede garantizar que la solución podrá ser usada de forma eficiente a través de dispositivos móviles?

Para crear una solución a este problema se propone un modelo de índice invertido más bolsa de palabras. Con este modelo es necesario crear un vocabulario con multitud de palabras de todas las imágenes de los edificios históricos de Granada. Para crear el vocabulario se recorren todas las imágenes y, por cada una, se guardan los descriptores de las zonas más representativas.

El siguiente paso es crear un histograma para cada imagen utilizando las palabras del vocabulario y los descriptores de dicha imagen.

Cuando se tienen los histogramas de las imágenes, si un usuario quiere obtener información sobre un edificio, debería realizar una foto sobre él. A la foto que ha realizado el usuario se calcula su histograma y se compara con todos los que se encuentran en la base de datos. Cuando termine el proceso se pueden ordenar los histogramas por la distancia entre ellos y se elige el que menor distancia tiene. Para obtener la información sobre el edificio, se relaciona cada imagen de la base de datos con su correspondiente información, lo que permite devolverla al usuario.

Para garantizar que la solución se puede usar de forma eficiente en los dispositivos móviles, la aplicación podría tener ya establecido el sistema con todos los histogramas y bolsa de palabras calculados, por lo que solo debería calcular el histograma de la imagen sobre la que busca información y hacer la comparación. Otra opción más rápida para el usuario consiste en enviar a un servidor (con gran capacidad de cómputo) la imagen sobre la que se quiere realizar la consulta para que sea éste el que calcule todas las operaciones y devuelva solo la información.

8. Suponga que desea detectar la presencia/ausencia de señales de tráfico en imágenes tomadas desde una cámara situada en la parte frontal de un coche que viaja por una carretera. Diga qué aproximación usaría y por qué. Identifique las principales dificultades y diga cómo las resolvería. Los argumentos deben ser sólidos y con fundamento en las técnicas estudiadas.

Para detectar la presencia/ausencia de señales de tráfico utilizaría la técnica "Generalized Hough Transform". Los motivos por los que elegiría esta técnica son que permite encontrar formas parciales o deformadas, por lo que si una señal se ha partido o se ha doblado por cualquier motivo podría encontrarla. Otro motivo es que puede encontrar estructuras adicionales en la imagen, que permitiría reconocer la señal si ha sido pintada, parcialmente tapada, etc. Un motivo importante es la posibilidad de encontrar varios objetos en una misma imagen, por lo que, si existen varias señales en el mismo momento, las podrá reconocer todas.

Las principales dificultades residen en el hardware del dispositivo que es necesario para realizar los cálculos, ya que requiere grandes requerimientos de cómputo y almacenamiento. Al tratarse de información que se debe mostrar en tiempo real, la solución sería instalar dispositivos más potentes, buscando un equilibrio entre precio del dispositivo y el rendimiento deseado.

9. ¿Qué han aportado los modelos CNN respecto de los modelos de reconocimiento de objetos empleados hasta 2012? Enumerar las propiedades comunes entre ellos y aquéllas claramente distintas que hayan permitido una mejora en la solución del problema por parte de las CNN. Dar una opinión razonada de por qué significan realmente una mejora.

Los modelos CNN convolucionan las características aprendidas con las imágenes de entrada, utilizando capas en dos dimensiones que hacen que sean muy buenas para el tratamiento de imágenes.

Lo que han aportado los modelos CNN respecto de los modelos de reconocimiento de objetos estudiados es que no es necesario identificar las características que se deben utilizar para clasificar las imágenes.

Las propiedades que comparten los modelos es la propagación hacia atrás, aunque varía la aplicación entre capas, ya que en las primeras versiones se tenían capas muy profundas que hacían que no funcionaran correctamente en problemas complejos. El cambio que provocaron los modelos CNN reside en cambiar este enfoque, creando capas más anchas que permitían solucionar el problema con los casos complejos.

10. Razone y argumente a favor y en contra de usar modelos de redes CNN ya entrenados y que se conocen han sido efectivos en otras tareas distintas de la que tiene que resolver, como modelos para aplicar directamente o como modelos a refinar para la tarea que tiene entre manos. Dar argumentos que no sean genéricos o triviales y que fundamenten su postura.

Usar modelos de redes CNN ya entrenados pueden aportar grandes resultados (como se comentó en clase sobre un modelo CNN de Google ya entrenado que se utilizó para detectar tumores en imágenes). Si las capas iniciales se encargan de aprender características de imágenes que van recibiendo, sería necesario simplemente ajustar los parámetros en las capas finales para que permitan clasificar correctamente los datos de cada problema concreto.

11. Referencias:

- Tema 5. Feature Matching.
- Tema 11. Introduction to Recognition.
- Tema 13. Object instance recognition.
- Learning Features: Convolutional Neural Networks (CNN).
- https://es.wikipedia.org/wiki/Modelo_bolsa_de_palabras
- https://en.wikipedia.org/wiki/Generalised_Hough_transform
- https://es.mathworks.com/discovery/deep-learning.html