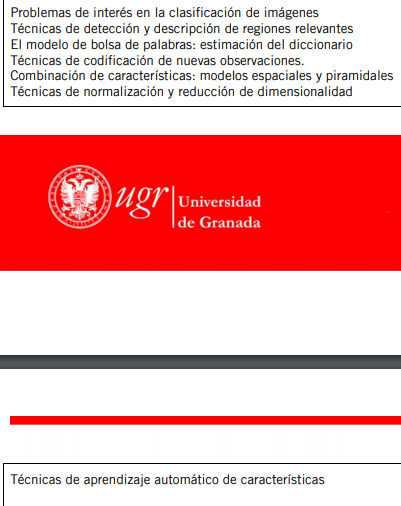
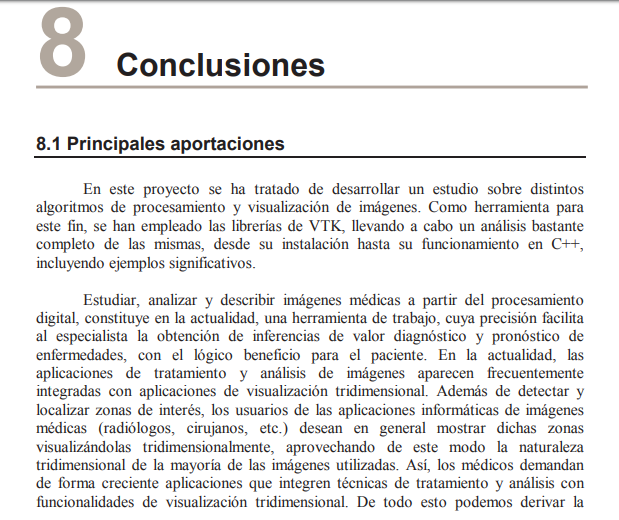


Problemas de interés en la clasificación de imágenes Técnicas de detección y descripción de regiones relevantes El modelo de bolsa de palabras: estimación del diccionario Técnicas de codificación de nuevas observaciones. Combinación de características: modelos espaciales y piramidales Técnicas de normalización y reducción de dimensionalidad  

En primer lugar, tenemos que distinguir la clasificación de la detección de objetos.

En la detección, lo que queremos hacer es localizar en una imagen,

todas las apariciones de un determinado objeto.

En este ejemplo, personas.

La detección va a a implicar encontrar dónde está cada objeto,

mientras que en la clasificación únicamente asignamos una

etiqueta a toda la imagen, sin importarnos dónde está el objeto.

Decir que en Coursera existe otro curso específico sobre detección,

en el que también hemos participado algunos de los profesores que impartimos

este curso de clasificación.

Tenemos también que distinguir la clasificación de la segmentación.

En la segmentación, el objetivo es descomponer la imagen en los diferentes

objetos o elementos que la componen.

Por ejemplo en este caso, en esta imagen tenemos el cielo el puente,

la persona en bicicleta, la ciudad, el río, etcétera.

Y finalmente, hay que distinguir la clasificación de lo que llamamos

búsqueda por similitud, o retrieval en inglés,

donde el objetivo va a ser buscar dentro de una base de datos de imágenes, otras

imágenes que sean visualmente similares a la imagen original que tenemos.

:\_\_\_\_\_\_\_\_\_\_\_---

vamos a ver por qué es difícil realizar la clasificación.

La principal dificultad vendrá dada por la enorme variabilidad que podemos tener

en la apariencia visual de imágenes de la misma clase.

Esta variabilidad puede venir dada por multitud de factores.

Por ejemplo, diferencias en la iluminación que tenemos en la imagen, en el tamaño,

en la posición, o en el punto de vista del objeto principal,

aquí tenemos caballos que son muy pequeños, caballos que son más grandes,

caballos que están de espaldas, caballos que están de lado.

También podemos tener variabilidad en el fondo de la imagen,

aquí tenemos un caballo en una montaña, en una playa, en un prado, etcétera.

También podemos tener oclusiones del objeto a reconocer, en este caso los

caballos están tapados por una persona, aquí está medio escondido dentro del agua,

o bien, la propia variabilidad del objeto que queremos identificar,

ya que tenemos diferentes tipos o razas de caballos.

Otro motivo de dificultad va a ser también, el gran número de etiquetas o

de categorías que podemos asignar en una, a una imagen.

En el diccionario, podemos encontrar del orden de miles de palabras que son

susceptibles de poder ser utilizadas como categorías en una imagen.

 :\_\_\_\_\_\_\_\_\_\_\_\_--

Normalmente, los sistemas que diseñaremos nos permitirán distinguir

solo entre un número más reducido de categorías, del orden de las decenas,

de las centenas, incluso de unos pocos miles.

 \_\_\_\_\_\_\_

A lo largo del curso seguiremos este esquema general que vemos aquí

en el que tenemos tres pasos bien diferenciados.

Por un lado la extracción de características,

la representación de la imagen y la clasificación.

El objetivo va a ser que a partir de una imagen de entrada,

al final podamos asignarle una determinada categoría,

de entre las que habremos aprendido a partir del conjunto de aprendizaje.

 \_\_\_\_\_\_\_\_\_--

INICIO DE COMO CLASIFICAR

Vamos a introducir ahora brevemente cada uno de estos tres pasos.

El primer paso va a ser siempre lo que llamaremos extracción de características,

que nos va a permitir obtener

características visuales a partir de la imagen.

7:53

Veremos que cada característica visual va a acabar siendo un vector numérico que nos

describirá de forma matemática la apariencia visual

de una región de la imagen.

En este punto permitidme hacer una pequeña aclaración

para aquellos de vosotros que hayáis seguido el curso de detección de objetos.

8:10

Para la clasificación,

y debido a la gran variabilidad que podemos tener que ya hemos comentado,

este enfoque nos introduciría mucho ruido y variabilidad, con lo que utilizaremos

una técnica diferente que consiste en dos procesos diferentes.

 En primer lugar vamos a detectar en la imagen lo que llamaremos puntos de

interés, es decir,

zonas que nos puedan dar una información relevante sobre el contenido de la imagen.

8:54

Normalmente, estos puntos estarán localizados en zonas significativas,

como por ejemplo zonas con cambios significativos de contraste,

o bien en los contornos de los objetos.

En segundo lugar, para cada uno de estos puntos de interés,

vamos a obtener una descripción local.

Es decir, una descripción del contenido visual en una pequeña

zona de la imagen que definiremos alrededor del punto de interés.

Esta descripción va a ser un vector numérico que, como hemos dicho

nos va a dar una representación matemática del contenido visual de esa zona.

\_\_\_\_\_\_\_\_\_\_\_\_\_---

Así pues, resumiendo este primer paso de extracción de características,

hemos visto que se compone de dos subprocesos, la detección

y la descripción de puntos de interés, y como resultado nos da un conjunto de

puntos de interés en la imagen, que podrá ser variable de una imagen a otra,

y una descripción en forma de vector numérico, para cada uno de estos puntos.

9:51

Esta lista con la descripción de todos los puntos de interés

nos da la información básica sobre el contenido visual de las imágenes,

pero es difícil de manejar desde el punto de vista computacional,

tanto por la cantidad de información como por el formato, ya que,

tal como veremos, los mejores clasificadores están diseñados para

\_\_\_\_\_\_\_\_\_\_\_\_-----

trabajar a partir de una representación en forma de un único vector numérico.

Por lo tanto, el segundo paso en nuestro esquema será obtener una representación

única de toda la imagen, en forma de vector numérico.

Esta representación la vamos a obtener agregando o combinando

la descripción de todos los puntos de interés.

\_\_\_\_\_\_\_\_\_---

A partir de esta representación única de toda la imagen,

el paso final va a ser el paso de clasificación propiamente dicho.

Es en este último paso donde realmente vamos a aprender,

a partir del conjunto de aprendizaje, cómo distinguir las imágenes de las

diferentes categorías a partir de la representación visual de cada imagen.

Es decir, vamos a tener,

para cada imagen, una representación en un espacio multidisciplinar.

Para simplificar, en el, y para dibujarlo, vamos a suponer que tenemos un espacio

bidimensional, por lo tanto, vectores de dos dimensiones.

Por lo tanto, esta imagen la podemos representar en este eje de coordenadas.

Podemos hacer lo mismo para la, eh, cada una de las imágenes que

tenemos en las otras clases, y para todas las imágenes del conjunto de aprendizaje.

11:21

De esta forma,

vamos a tener todas las imágenes del conjunto de aprendizaje representadas en

el espacio multidimensional que nos da la representación de la imagen.

11:31

Si esta representación es suficientemente discriminativa, como esperamos,

nos deberíamos encontrar que las imágenes de las diferentes clases deberían quedar

representadas en zonas diferentes de este espacio multidimensional.

\_\_\_\_\_\_\_\_\_\_---

12:14

A lo largo del curso iremos explorando diferentes alternativas para cada uno de

estos pasos del esquema general que vamos a introducir.

En la primera semana veremos un sistema básico que cubrirá todos los es,

todos los pasos del esquema, de una forma simple.

En la segunda semana introduciremos una nueva forma de obtener la

representación global de la imagen, es lo que se conoce como bag of words en inglés.

Este método de representación ha sido la base estándar de representación

de imágenes para clasificación a lo largo de los últimos 10 años,

y será la representación sobre la que basaremos la explicación

de la mayoría del contenido del curso.

----------------\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

métodos de clasificación más utilizados,

que se conoce como support vector machine, en inglés, y que también será el método

que utilizaremos mayoritariamente a lo largo de todo el curso.

13:05

En la tercera semana, exploraremos diferentes alternativas para realizar

el primer paso de detección y descripción de puntos de interés, incorporando también

descriptores que permiten incorporar diferentes tipos de contenido visual,

como por ejemplo la forma y el color.

\_\_\_\_\_\_\_-----

RESUMEN

 puntos de interés de la imagen, características que nos interesará

que sean repetitivas en varias imágenes y que además sean discriminativas.

por el concepto mismo de imagen, en general vamos a trabajar

con imágenes en color que estarán capturadas por cámaras digitales estándar.

El resultado que se obtiene cuándo capturamos una imagen con una cámara

depende esencialmente de tres factores.

En primer lugar el color de la luz que ilumina la escena,

en segundo lugar el tipo de material de la escena que estamos capturando y que

tendrá unas propiedades de la refracción de la luz diferentes para cada material

y por último la sensibilidad de la cámara que estamos utilizando que

evidentemente será diferente en función del modelo y de la calidad de la cámara.

\*\*\*-\*\*\*

Hay que decir que normalmente las cámaras tienen tres sensores, que son sensibles

a diferentes longitudes de onda de la luz y que habitualmente estas longitudes

a las que son sensibles se corresponden con los colores rojo, verde y azul.

Es con estos tres sensores que las cámaras intentan cubrir la sensibilidad

a todo el espectro visible.

De esta forma, el resultado final cuando capturamos una imagen es una matriz de

puntos, lo que normalmente conocemos como los píxeles de la imagen.

Si ampliamos una eternidad la zona de esta imagen

podemos ver que el valor del color con cada uno de los píxeles de la imagen se

obtiene como la combinación de tres valores.

Que corresponden al canal rojo, al canal verde y al canal azul,

RGB de las siglas en inglés de red, green and blue.

Por lo tanto cada píxel viene dado por tres valores numéricos, que en general

están en el rango entre 0 y 255, donde el 0 indica la ausencia de color

y el 255 indica la máxima representación de ese color en un punto.

\_\_\_\_\_\_\_---

CONCLUSION

Comentemos también, que el detector se complementa con una segunda parte

que es un descriptor de características y que veremos en el próximo vídeo.

De esta forma, podemos no solo detectar los puntos de interés, si no también,

representar la información que tenemos en la imagen alrededor de estos puntos.

Esta información, está descripción, es la que utilizaremos despues para representar

y para clasificar las imágenes.

2:00

\_\_\_\_\_\_\_\_\_-