# Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación Tratamiento de Señales Visuales (TSV)

Grado en Ingeniería Informática Introducción a la Visión Artificial (IVA)

# Práctica 3 Reconocimiento de Escenas con modelos Bag-of-Words









### **OBJETIVO**



 Reconocimiento de escenas mediante el uso de descriptores y clasificadores basados en bag of words



Imagen test



## **ENTREGA**



- Descargue los ficheros de código Python disponibles en Moodle
  - -Fichero p3 bow code.zip
- Complete las tareas utilizando los ficheros \*.py contenidos en el ZIP
- La entrega se realiza por parejas y constará de:
  - -Memoria en PDF (1 fichero)
    - + código Python adicional
  - -Ficheros código Python de las tareas realizadas (hasta 2 ficheros)
  - -Fichero p3\_tests.pyc y p3\_utils.pyc
  - –No entregar carpetas 'dataset' y 'test\_data'



No es necesario entregar todas las tareas



#### **TAREAS**



- Tarea 0 Descargar material e incluir nombre en ficheros
- Tarea 1 Implementar el modelo BOW
- Tarea 2 Implementar descriptores Tiny/HOG
- Tarea 3 Responder a las preguntas en memoria
- Funciones prohibidas:
  - -Cualquier función que no sea del paquete de numpy y scipy
  - -Se pueden utilizar las funciones recomendadas de skimage
  - -Para normalización de imágenes, no se pueden utilizar las funciones img\_as\_float64, img\_as\_float32, y img\_as\_float.

    Se debe realizar la operación manualmente.

Aumento de complejidad con respecto a la práctica anterior y mayor número de preguntas en memoria



## TAREAS – IMPORTANTE!



 En cada fichero debe poner la ruta donde se encuentra el dataset que ha descargado de Moodle

```
    p3_tarea2.py 1 • • p3_tarea1.py 1 ×

p3_2122 > 🗣 p3_tarea1.py > 🕅 obtener_bags_of_words
                                                                                                                              C:\practicasTSV\datasets\scenes15
          vocabulario = np.empty(shape=[vocab_size,list_img_desc[0].shape[1]]) # iniciamos la variable de salida
                                                                                                                                 Nombre
                                                                                                                                     Bedroom
         return vocabulario
      def obtener_bags_of_words(list_img_desc, vocab):
                                                                                                                                     Coast
         # Esta funcion obtiene el Histograma Bag of Words para cada imagen
                                                                                                                                     Forest
         # Argumentos de entrada:
                                                                                                                                     Highway
         # - list_img_desc: Lista 1xN con los descriptores de cada imagen. Cada posicion de la lista
                            contiene (MxD) numpy arrays que representan UNO O VARIOS DESCRIPTORES
                                                                                                                                     Industrial
                             extraidos sobre la imagen
                             - M es el numero de vectores de caracteristicas/features de cada imagen
                                                                                                                                     InsideCity
                             - D el numero de dimensiones del vector de caracteristicas/feature.
             - vocab: Numpy array de tamaño [vocab_size, D],
                                                                                                                                     Kitchen
                                                                                                                                     LivingRoom
         # Argumentos de salida:
             - list_img_bow: Array de Numpy [N x vocab_size], donde cada posicion contiene
                             el histograma bag-of-words construido para cada imagen.
                                                                                                                                     Mountain
                                                                                                                                     Office
         list_img_bow = np.empty(shape=[len(list_img_desc),len(vocab)])
                                                                                                                                     OpenCountry
                                                                                                                                     Store
         return list_img_how
                                                                                                                                     Street
     if __name__ == "__main__":
          dataset_path = './datasets/scenes15/'
                                                                                                                                     Suburb
         print("Practica 3 - Tarea 1 - Test autoevaluación\n")
         print("Tests completados = " + str(test_p3_tarea1(dataset_path,stop_at_error=False,debug=False))) #ana1
                                                                                                                                     TallBuilding
```

dataset\_path = 'C:/practicasTSV/datasets/scenes15/'

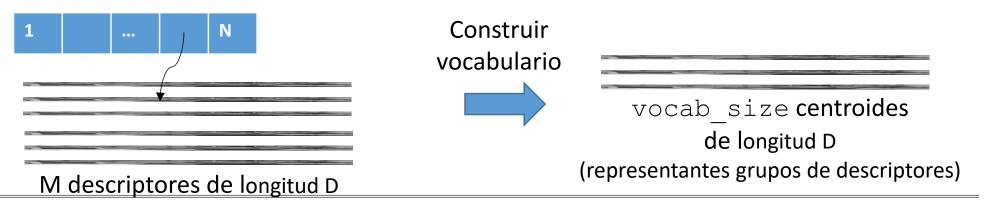


# TAREA 1 1/2



- Implementar el modelo Bag-Of-Words (p3\_tarea1.py)
  - -Función construir\_vocabulario(list\_img\_desc, vocab\_size=10, max\_iter=300)
  - -Consideraciones:
    - Es resultado es el vocabulario en un array numpy con dimensión [vocab\_size,D]
  - -Funciones recomendadas:
    - Para agrupar, utilice la función sklearn.cluster.Kmeans con el parámetro random\_state=0. En esta función utilice los parámetros vocab size y max iter de la función construir\_vocabulario
    - Para convertir la lista en arrays numpy, utilice numpy.concatenate

#### Lista con descriptores para N imágenes

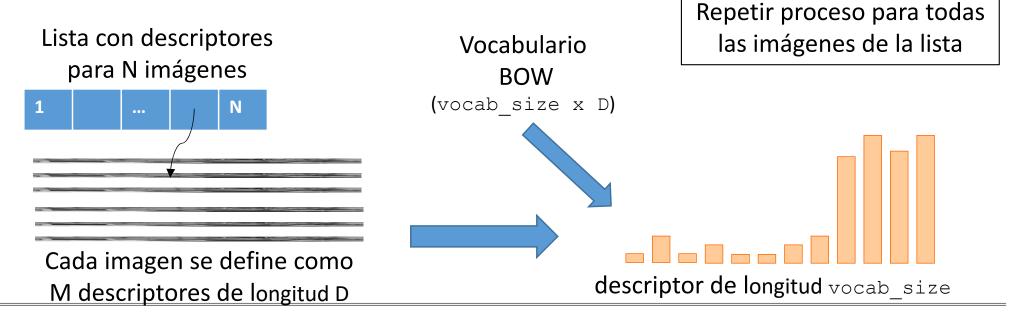




# TAREA 1 2/2



- Implementar el modelo Bag-Of-Words (p3\_tarea1.py)
  - -Función obtener\_bag\_of\_words(list\_img\_desc, vocab)
  - -Consideraciones:
    - El resultado es un array numpy con dimensión [N, vocab\_size] donde cada fila es el histograma BOW (normalizado) de cada imagen
  - -Funciones recomendadas:
    - Para calcular distancias utilice scipy.spatial.distance.cdist con valores por defecto.



# TAREA 2 1/2



## • Implementar extracción de descriptores (p3\_tarea2.py)

- -Función obtener\_features\_tiny(path\_imagenes,tamano=16)
- -Consideraciones:
  - Convierta cada imagen a gris, formato float y en el rango [0,1]
  - La salida de esta función es una lista 1xN, donde cada posición es un numpy array con los descriptores calculados para cada imagen

#### -Funciones

- Recomendadas: numpy.reshape, skimage.io.imread, skimage.color.rgb2gray, skimage.transform.resize
- Prohibidas: Cualquiera que no sea de numpy ó scipy

Repetir proceso para todas las imágenes de la lista

Redimensión a tamano X tamano

Vectorización a 1 X tamano<sup>2</sup>

Matriz MxD













- Implementar extracción de descriptores (p3\_tarea2.py)
  - -Función obtener\_features\_hog(path\_imagenes,tamano=100, orientaciones=9, pixeles\_por\_celda=(8, 8),celdas\_bloque=(4,4)
  - -Consideraciones:
    - Convierta cada imagen a gris, formato float y en el rango [0,1]
    - La salida de esta función es una lista 1xN, donde cada posición es un numpy array con los descriptores calculados para cada imagen

#### -Funciones

- Para calcular HOG, utilice skimage.feature.hog con los parámetros orientations=orientaciones, pixels\_per\_cell=pixeles\_por\_celda, cells\_per\_block=celdas\_bloque, feature\_vector=False (resto por defecto)
- Recomendadas (Skimage): io.imread, color.rgb2gray, transform.resize

Repetir proceso para todas las imágenes de la lista

Redimensión a Cálculo Matriz MxD
tamano X tamano HOG





## TAREA - MEMORIA



- Responda a las preguntas en un fichero "preguntas\_P3.docx"
  - No necesita realizar funciones nuevas, el trabajo se basa en realizar experimentos con la funcionalidad que ha generado
  - Se sugiere emplear búsqueda en grid para encontrar parámetros óptimos en cada pregunta.
  - -Puede necesitar repasar algunos conceptos teóricos
  - -Utilice figuras y tablas para resumir experimentos
- Ficheros a generar
  - -Fichero "preguntas\_P3.pdf" en formato PDF con las respuestas
  - -Genere tantos ficheros Python como necesite con el formato p3\_pregunta\_XX.py donde XX es el número de pregunta
- Criterios generales de evaluación: texto conciso y claro; relación con teoría; numeración de diagramas, figuras y tablas; referencias a figuras y tablas en texto; referencias a fuentes externas (si procede); errores ortográficos.
- Seis (6) preguntas en total agrupadas por temática.



# TAREA 3 – MEMORIA 1/3



#### P3.1 Cree un esquema de clasificación de imágenes con los siguientes detalles:

- + Características: HOG con parámetro tam=100
- + Modelo: BOW con max\_iter=10
- + Clasificador: KNN con K=5
- + Otros: Ratio train\_test=0.20, y Máx número datos por categoría: 200

Debe utilizar las funciones para obtener features en p3\_tarea2.py, funciones construir\_vocabulario/obtener\_bags\_of\_words (fichero p3\_tarea1.py), y función load\_image\_dataset (fichero p3\_utils.pyc). Para construir este esquema puede basarse en sklearn.model\_selection.train\_test\_Split, sklearn.neighbors.KNeighborsClassifier

Y aplíquelo sobre el dataset scene15 disponible en el material de la práctica con una partición determinista. A continuación, responda a las siguientes preguntas:

- 3.1.1 Compare los resultados obtenidos al utilizar las características HOG y Tiny (con parámetro tam=64), en términos de rendimiento de clasificación sobre los datos de entrenamiento y test. Utilice valores por defecto indicados en el enunciado con un tamaño de diccionario de 50. (0.75 puntos)
- 3.1.2 Varíe el tamaño del diccionario BOW (hasta un valor máximo de 200) y estudie el rendimiento de clasificación sobre los datos de entrenamiento y test. Utilice valores por defecto indicados en el enunciado. (0.75 puntos)
- 3.1.3 Varíe el número de vecinos del clasificador KNN (hasta 21) con HOG y estudie el rendimiento de clasificación sobre los datos de entrenamiento y test. Muestre resultados visuales de aciertos/errores (se recomienda la función create\_webpage\_results). Utilice valores por defecto indicados en el enunciado con el tamaño de diccionario BOW óptimo obtenido en la pregunta 3.1.2. (1.0 puntos)



# TAREA 3 – MEMORIA 2/3



#### P3.2 Cree un esquema de clasificación de imágenes con los siguientes detalles:

- + Características: HOG con parámetro tam=100
- + Modelo: BOW con max\_iter=10
- + Clasificador: SVM (lineal)
- + Otros: Ratio train\_test=0.20, y Máx número datos por categoría: 200

Debe utilizar las funciones para obtener features en p3\_tarea2.py, funciones construir\_vocabulario/obtener\_bags\_of\_words (fichero p3\_tarea1.py), y función load\_image\_dataset (fichero p3\_utils.pyc). Para construir este esquema puede basarse en sklearn.model\_selection.train\_test\_Split, sklearn.svm

Y aplíquelo sobre el dataset scene15 disponible en el material de la práctica con una partición determinista. A continuación, responda a las dos siguientes preguntas:

- 3.2.1 Varíe el tamaño del diccionario BOW (hasta un valor máximo de 200) y estudie como varía el rendimiento de clasificación sobre los datos de entrenamiento y test. Utilice valores por defecto indicados en el enunciado. (1.0 puntos)
- 3.2.2 Investigue y compare los distintos tipos de kernels no lineales para clasificadores SVM (e.g., rbf y poly) sobre los datos de entrenamiento y test. Razone por-que se obtiene unos resultados distintos a la pregunta anterior con una SVM lineal. Posteriormente, analice el rendimiento de la configuración con el mejor resultado y muestre resultados visuales de aciertos/errores (se recomienda la función crea-te\_webpage\_results). Utilice valores por defecto indicados en el enunciado con el tamaño de diccionario BOW óptimo obtenido en la pregunta 3.2.1. (1.25 puntos)



# TAREA 3 – MEMORIA 3/3



#### P3.3 Cree un esquema de clasificación de imágenes con los siguientes detalles:

- + Características: HOG con parámetro tam=100
- + Modelo: BOW con max\_iter=10
- + Clasificador: Random Forest
- + Otros: Ratio train\_test=0.20, y Máx número datos por categoría: 200

Debe utilizar las funciones para obtener features en p3\_tarea2.py, funciones construir\_vocabulario/obtener\_bags\_of\_words (fichero p3\_tarea1.py), y función load\_image\_dataset (fichero p3\_utils.pyc). Para construir este esquema puede basarse en sklearn.model\_selection.train\_test\_Split, sklearn.ensemble.RandomForestClassifier

Y aplíquelo sobre el dataset scene15 disponible en el material de la práctica con una partición determinista. A continuación, responda a las dos siguientes preguntas:

3.3.1 Investigue la función sklearn.ensemble.RandomForestClassifier y seleccione solo uno de los parámetros disponibles. Compare y razone los resultados para distintos valores del parámetro seleccionado. Posteriormente, analice el rendimiento de la configuración con el mejor resultado y muestre resultados visuales de aciertos/errores (se recomienda la función create\_webpage\_results). Utilice valores por defecto indicados en el enunciado con el tamaño de diccionario BOW óptimo obtenido en la pregunta 3.2.1. (1.25 puntos)



# TAREA 3 – MEMORIA – FUNCIONES AYUDA



- Las preguntas requieren construir varios sistemas de clasificación del dataset con tareas 1/2, clasificadores KNN/SVM/RF y características
  - -Opción 1: Raw/descriptores
  - -Opción 2: **BOW** descriptores

Lectura de dataset División de datos entre train/test Obtención características (e.g. HOG) Entrenar clasificador con datos *train* Evaluar clasificador

con datos test

Lectura de dataset

División de datos entre train/test

Obtención características (e.g. HOG)

> Vocabulario BOW con datos train

Obtener descriptores BOW para train/test

Entrenar clasificador con datos *train* 

Evaluar clasificador con datos test



# **UAM** TAREA 3 – MEMORIA – FUNCIONES AYUDA



• Funciones útiles: lectura del dataset (fichero p3 utils.pyc) 1/2

```
data = load_image_dataset(container_path, *, description=None,
categories=None,load_content=True, shuffle=True,random_state=0,
resize_shape=None, max_per_category=None,debug=False):
```

Datos de imágenes

Etiquetas

Etiquetas (nombres)

Rutas de imágenes

data : :class:`~sklearn.utils.Bunch` Dictionary-like object, with the following attributes.

data: list of str

Only present when `load content=True`.

The raw image data to learn.

target : ndarray

The target labels (integer index).

target names : list

The names of target classes.

DESCR : str

The full description of the dataset.

filenames: ndarray

The filenames holding the dataset



# TAREA 3 – MEMORIA – FUNCIONES AYUDA



• Funciones útiles: lectura del dataset (fichero p3 utils.pyc) 2/2

```
data = load_image_dataset(container_path, *, description=None,
categories=None,load_content=True, shuffle=True,random_state=0,
resize_shape=None, max_per_category=None,debug=False):
```

```
#example 1 - Load dataset
# 100 images per category
# resize all images to (200,200))
dataset path = './datasets/scenes15'
data = load_image_dataset(container_path=dataset_path, shuffle=True
        load_content=True, resize_shape=(200,200),max_per_category=100)
```

```
#example 2 - Load dataset
# 200 images per category
# does not load images (only paths) - useful for computing features later
dataset path = './datasets/scenes15'
data = load image dataset(container path=dataset path, shuffle=True,
               load content=False, max per category=200)
```



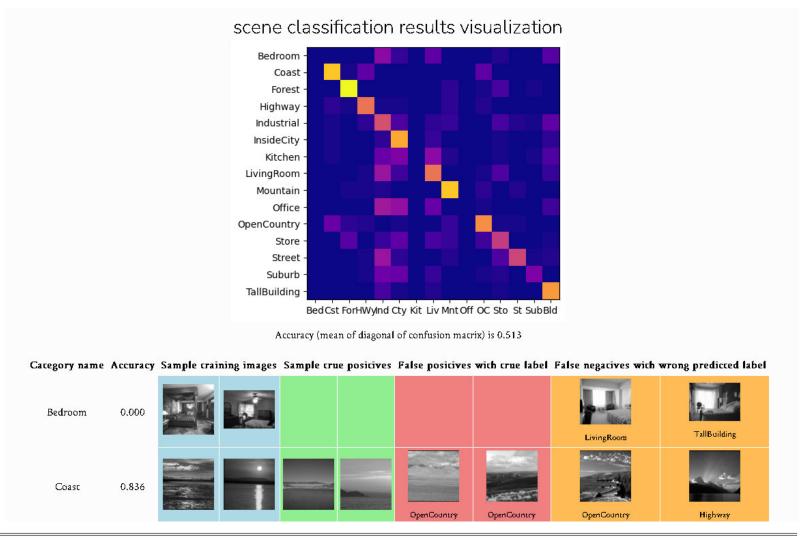
# TAREA 3 – MEMORIA – FUNCIONES AYUDA



• Funciones útiles: visualización resultados (fichero p3 utils.pyc) 1/3

create\_results\_webpage(...):

La función genera una página web html con resultados cualitativos y cuantitativos





# **UAM** TAREA 3 – MEMORIA – FUNCIONES AYUDA



• Funciones útiles: visualización resultados (fichero p3 utils.pyc) 2/3

```
confusion_matrix = create_results_webpage
(train_image_paths, test_image_paths,
train_labels, test_labels, categories, abbr_categories, predic
ted categories, name experiment):
```

- confusion matrix → matriz de confusión obtenida
- train image paths → Lista de Python con las rutas (relativas o absolutas) de las imágenes de entrenamiento

test\_image\_paths → Lista de Python con las rutas (relativas o absolutas) de las

imágenes de entrenamiento

```
-train labels,
```

- -test labels
- -predicted\_categories

Lista de Python con las categorías reales para los datos de train/test y predichas solamente para los datos de test

```
['TallBuilding', 'TallBuilding', 'Kitchen', 'Bedro...
> special variables
> function variables
  0000: 'TallBuilding'
        'TallBuilding'
         'InsideCity'
 0006: 'Kitchen'
 0007: 'InsideCity'
        'TallBuilding'
        'Highway'
```



# **UAM** TAREA 3 – MEMORIA – FUNCIONES AYUDA



• Funciones útiles: visualización resultados (fichero p3 utils.pyc)

3/3

```
confusion matrix = create_results_webpage
(train_image_paths, test_image_paths,
train_labels, test_labels, categories, abbr_categories, predicted
_categories, name_experiment):
# This is the list of categories / directories to use.
# The categories are sorted by alphabetical order
categories = ['Bedroom', 'Coast', 'Forest', 'Highway', 'Industrial',
              'InsideCity', 'Kitchen', 'LivingRoom', 'Mountain', 'Office',
              'OpenCountry', 'Store', 'Street', 'Suburb', 'TallBuilding']
# This list of shortened category names is used later for visualization.
abbr categories = ['Bed', 'Cst', 'For', 'HWy', 'Ind', 'Cty', 'Kit', 'Liv', 'Mnt',
                'Off', 'OC', 'Sto', 'St', 'Sub', 'Bld']
```

name\_experiment → string que describe el experimento realizado. Los resultados se guardan en la carpeta ./p3\_results\_webpage/ name\_experiment (e.g. name\_experiment = 'HOG\_BOW\_SVM', genera la página html de resultados en ./p3\_results\_webpage/ HOG\_BOW\_SVM/index.html)



# TAREA 3 – MEMORIA – FUNCIONES AYUDA



#### • Funciones útiles:

- -Para dividir el dataset entre train/test (paquete sklearn)
  - sklearn.model\_selection.train\_test\_split
- -Para clasificar, KNN y SVM del paquete sklearn
  - KNN sklearn.neighbors.KNeighborsClassifier
  - **SVM** sklearn.svm
  - RF sklearn.ensemble.RandomForestClassifier
- -Para evaluar el rendimiento el dataset
  - sklearn.neighbors.KNeighborsClassifier.score
  - sklearn.svm.score
  - sklearn.ensemble.RandomForestClassifier.score

#### –Documentación:

- https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html#sklearn.svm.SVC
- https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html
- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html



## TAREAS – AUTOEVALUACION



- ¿Cómo comprobar si la implementación es correcta?
  - -Se proporciona funciones que se pueden ejecutar cuando se desee
    - test\_p3\_tarea1(...) en el fichero p3 tarea1.py

```
Practica 3 - Tarea 1
Realizando tests para las funciones 'construir_vocabulario' y 'obtener_bags_of_words' de la tarea 1
La funcion es correcta si los resultados obtenidos tienen una tolerancia de 2 decimales con respecto a la salida correcta.
 * Utilizando datos en fichero '.\p3_tarea1.data'
 * Tests para 'construir_vocabulario':
        Testeando con tamano de vocabulario 5 (20 descriptores Tiny de dimension 256)... OK
        Testeando con tamano de vocabulario 5 (1620 descriptores HOG de dimension 144)... OK
        Testeando con tamano de vocabulario 10 (20 descriptores Tiny de dimension 256)... OK
        Testeando con tamano de vocabulario 10 (1620 descriptores HOG de dimension 144)... OK
 * Tests para 'obtener_bags_of_words':
        Testeando con tamano de vocabulario 5 (20 descriptores tipo Tiny de dimension 256)... OK
        Testeando con tamano de vocabulario 5 (1620 descriptores tipo HOG de dimension 144)... OK
        Testeando con tamano de vocabulario 10 (20 descriptores tipo Tiny de dimension 256)... OK
        Testeando con tamano de vocabulario 10 (1620 descriptores tipo HOG de dimension 144)... OK
 * Finalizado en 2.435 secs
 * RESULTADO 'construir_vocabulario' con descriptor Tiny/HOG: 4/4 CORRECTOS ( 100.00% )
 * RESULTADO 'obtener_bags_of_words' con descriptor Tiny/HOG: 4/4 CORRECTOS ( 100.00% )
Tests completados = True
```

test\_p3\_tarea2(...) en el fichero p3 tarea2.py

```
Practica 3 - Tarea 2
Realizando tests para las funciones 'obtener_features_tiny' y 'obtener_features_hog' de la tarea 2
La funcion es correcta si los resultados obtenidos tienen una tolerancia de 2 decimales con respecto a la salida correcta.

* Utilizando datos en fichero '.\p3_tarea2.data'

* Tests para 'obtener_features_hog':
        Testeando con tamano 50... Imagen #0 #1 #2 #3 #4 #5 #6 #7 #8 #9
        Testeando con tamano 100... Imagen #0 #1 #2 #3 #4 #5 #6 #7 #8 #9

* Finalizado en 1.469 secs

* RESULTADO 'obtener_features_hog': 20/20 DESCRIPTORES CORRECTOS ( 100.00% )
Tests completados = True
```



## TAREAS – AUTOEVALUACION



No hay funciones de auto-evaluación para preguntas en la memoria pero...

- Accuracy/rendimiento de clasificación sobre datos de test (utilizando como entrenamiento 200 imágenes por categoría)
  - ~0% > funcionamiento erróneo
  - ~7% -> rendimiento similar a un clasificador aleatorio
  - ~18%  $\rightarrow$  rendimiento para características <u>Tiny con KNN (k=5)</u>
    - Se puede aumentar ligeramente ajustando parámetro k
  - ~40% → rendimiento para características HOG-BOW y KNN (k=5)
    Se puede aumentar ligeramente ajustando parámetro k
  - ~50% -> rendimiento para características HOG-BOW y SVM (lineal)
  - ~55% -> rendimiento para características HOG-BOW y RF
  - ~70% -> rendimiento para características HOG-BOW y SVM

Ajuste de parámetros SVM, tamaño de la imagen, extracción HOG y diccionario BOW.



## **CARGA DE TRABAJO**



- 2 sesiones de prácticas (20 horas): 4h presenciales + 16h no presenciales
- Sugerencia de realización:

TAREA	Horas presenciales	Horas no presenciales	Horas TOTAL	Semana de prácticas
Explicación	0.5 h	0 h	0.5 h	1
Tarea 1	1.5 h	3 h	4.5 h	1,2
Tarea 2	2 h	1 h	3 h	2
Memoria	0 h	12 h	12 h	2,3
TOTAL	4h	16h	20h	



## **EVALUACION**



• TAREA	\	Max nota	Criterio evaluado (ver rúbrica en Moodle)	
Tarea	1	2	Código: Ejecución (60%) + Diseño(40%)	
Tarea	2	2	Código: Ejecución (60%) + Diseño(40%)	
Memo	ria	6	Memoria: Claridad, exactitud de la respuesta y experimentos realizados por cada pregunta (100%)	
	P3.1	2.5		
	P3.2	2.25		
	P3.3	1.25		
TOTA	L	10		

#### Penalizaciones:

- Por entrega de ficheros no acorde a las especificaciones: -0.5 puntos
- Por entrega de memoria con formato incorrecto: -0.5 puntos
- Por uso de funciones prohibidas: -50% tarea
- Por entrega tardía (tras considerar los 4 días disponibles para cada pareja):
  - -25% (un día), -50% (dos días), -75% (tres días), -100%(>= días)



#### REFERENCIAS



## Wikipedia

https://en.wikipedia.org/wiki/Bag-of-words model

#### Descriptor Histograma de Gradientes Orientados

- Richard Szeliski, "Computer Vision: Algorithms and Applications"
   2020, Sección 7.1.2. <a href="https://szeliski.org/Book/">https://szeliski.org/Book/</a>
- -Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2), 91-110

## Modelo Bag-of-words

Richard Szeliski, "Computer Vision: Algorithms and Applications"
 2020, Sección 6.2. <a href="https://szeliski.org/Book/">https://szeliski.org/Book/</a>



#### **SUGERENCIA**



- Esta práctica presenta un incremento sustancial en complejidad respecto a prácticas anteriores.
- El trabajo efectivo y coordinado de la pareja es crítico para poder realizar todos los ejercicios satisfactoriamente.
- Sugerencia:
  - -Tareas 1 y 2: realizar conjuntamente
  - –Preguntas memoria:
    - Analizar que tareas son comunes a todas las preguntas (leer dataset, extraer características,...)
    - Repartir las tareas entre la pareja y realizar individualmente
    - Escribir los resultados en la memoria individualmente
    - Realizar un análisis y revisión conjunto de los resultados en memoria