Universidad de Chile
CC5509 Reconocimiento de Patrones
Prof. José M. Saavedra <jose.saavedra@orand.cl>

Tarea N° 3
Título: Reconocimiento de imágenes basado en *Bag of Words*

1. Objetivo: El objetivo de esta tarea es familiarizarse con la estrategia *Bag of Words* para el reconocimiento de imágenes. Particularmente, nos enfocaremos en el reconocimiento de perros. Esto permitirá, además, ganar experiencia con descriptores locales, *clustering* y clasificación.

2. Descripción: La tarea consiste de 3 etapas:
A) Generación de palabras visuales mediante *clustering*: En esta parte, se debe determinar un *corpus* formado por 5000 imágenes, bajo las cuales se determinarán *K* palabras visuales. Para obtener el corpus se deben seleccionar aleatoriamente 5000 imágenes del *dataset* Caltech-256 (http://www.vision.caltech.edu/Image_Datasets/Caltech256/).

Para cada imagen del *corpus* se debe calcular descriptores SIFT. Suponiendo que cada imagen genere aproximadamente 500 descriptores locales, en total tendremos 2.500.000 descriptores. Con el fin de facilitar el proceso se deben seleccionar aleatoriamente 100.000 descriptores que serán, finalmente, los que determinarán los *K* clústers. Luego, las *K* palabras visuales se obtienen mediante *clusterizar* el conjunto total de descriptores locales (los 100.000). Una palabra visual estará representada por el centroide de un determinado clúster. Para el proceso de *clustering* deben usar la estrategia K-MEANS.

Para obtener los descriptores locales puede usar diferentes libs como:

- VLFeat http://www.vlfeat.org/
- SIFT de OpenCV
- SIFT de Lowe (http://www.cs.ubc.ca/~lowe/keypoints/)
- Affine Covariant Features (http://www.robots.ox.ac.uk/~vgg/research/affine/index.html)

B) Entrenamiento usando SVN mediante histogramas *Bag of Words*: En esta etapa se debe modelar un clasificador que permita reconocer la ocurrencia ***de un perro*** en una imagen. Para este fin, es necesario obtener un **conjunto de entrenamiento**. Se recomienda obtener 500 imágenes de perros y 500 imágenes de no-perros. El dataset Caltech-256 contiene aproximadamente 100 imágenes de perros y miles de imágenes de no perros. Para completar el *dataset* de los patrones positivos (imágenes conteniendo un perro) se recomienda obtener las imágenes por *Google*. Para facilitar esta tarea, este *dataset* puede ser compartido entre todos los estudiantes del ramo.

Teniendo el conjunto de entrenamiento, cada imagen se representa como un histograma *Bag of Words*. Este histograma representa la distribución de ocurrencias de descriptores locales de la imagen entre los *K clústers* o palabras visuales. El histograma debe ser normalizado a la unidad.

La clasificación debe estar basada en la estrategia **Support Vector Machine.** Para este fin, se recomienda usar libSVM (http://www.csie.ntu.edu.tw/~cjlin/libsvm/). Deben experimentar con SVM + kernel RBF, los parámetros *C* y *gamma* deberán ser elegidos mediante validación cruzada. LibSVM incluye un pograma en Phyton para obtener los mejores parámetros.

***C)* Evaluación usando un conjunto de test**. Generar un conjunto de 200 imágenes de perros y 200 de no-perros. Este conjunto de imágenes no debe tener intersección con el *dataset* de entrenamiento.

La evaluación se debe representar mediante un gráfico de curva ROC definiendo un *threshold* relacionado al *score* de clasificación. Variando el *score* obtendrán diferentes tasas de clasificación. LibSVM permite modelar un SVM que produce un *score* entre 0 y 1 (modelo de probabilidad) si una imagen es de la clase positiva.

Además, deben experimentar con diferentes valores de *K (números de clústers)*. Presentar un gráfico que muestre la tasa de reconocimiento para diferentes valores de K, manteniendo fija la tasa de falsos positivos.

5. Informe: El informe debe escribirse en formato *paper* y debe incluir:

1) Introducción
2) Descripción del Trabajo
3) Evaluación y Análisis Resultados. Presentar ejemplos.
4) Conclusiones

Además, entregar el *dataset* de test generado!!!

6. Fecha de Entrega: 19, octubre, 2014. Las tareas deben ser enviadas por u-cursos.

Referencias

G. Csurka, C. Dance, L.X. Fan, J. Willamowski, and C. Bray (2004). "Visual categorization with bags of keypoints". Proc. of ECCV International Workshop on Statistical Learning in Computer Vision.

Sivic, J. and Zisserman, A. Video Google: A Text Retrieval Approach to Object Matching in Videos

Proceedings of the International Conference on Computer Vision (2003)