Pontificia Universidad Católica de Chile
IEE/IIC 3724: Reconocimiento de Patrones
Tarea 3: Recuperación de Imágenes por Contenido usando VLAD

Prof. José M. Saavedra
Ayudante: Erick Svec

Descripción General

En esta tarea se evaluará el desempeño de la estrategia VLAD (Vector of Locally Aggregated Descriptors) en el contexto de recuperación de imágenes por contenido (búsqueda por similitud). Para la evaluación se deberá utilizar el *dataset* Oxford (http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/).

Descripción Detallada

1) Descargar el *dataset* Oxford, el cual consta de 5062 imágenes, comprendiendo 11 diferentes categorías ("landmkarks").
   a) Dataset: http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/oxbuild_images.tgz
   b) Groundtruth: http://www.robots.ox.ac.uk/~vgg/data/oxbuildings/gt_files_170407.tgz
   c) Para cada categoría se proponen 5 consultas (imágenes de consultas), en total 55 consultas. Dada una consulta Q, cada una de las 5062 imágenes del *dataset* puede contener completamente a Q (etiqueta "Good"), contenerla parcialmente (etiqueta "OK" o "Junk") o no contenerla ("Bad"). Para este proyecto, siguiendo la metodología de evaluación de [1], solamente consideraremos como imágenes relevantes a Q, las imágenes etiquetadas como "Good" y como no relevantes, a aquellas etiquetadas como "Bad". La página web del *dataset* y el *paper* [1] contienen mayor descripción del *groundtruth*.
   d) Cuando descarguen el *dataset* observarán que éste contiene 5063 imágenes, esto se debe a que todas las imágenes de consulta están incluidas en el *dataset*. Así, cuando se realice la búsqueda por una consulta específica, tal consulta debe ser descartada del *dataset* (para no recuperarse a sí mismo) por lo que quedará un *dataset* con 5062 imágenes a recuperar.

2) Crear un **Codebook** usando la imágenes del *dataset* sin incluir a las imágenes de consultas (5063-55=5008 imágenes). Para este fin, utilice descriptores locales SIFT usando DoG como detector de puntos de interés. Aquí se recomienda utilizar **VlFeat** o la implementación de Krystian Mikolajczyk (http://kahlan.eps.surrey.ac.uk/featurespace/web/). **OpenCV** también es una muy buena alternativa.
   a) Para lo anterior se recomienda reducir el tamaño de las imágenes de modo que el ancho y alto no excedan 640 pixels, manteniendo la razón de aspecto. Con ellos debieran obtener aproximadamente 1000 descriptores por imagen (~5M en total), aunque la cantidad de descriptores depende también de la imagen por sí misma. Imágenes con más detalle producirán mas descriptores.
   b) Usando una muestra aleatoria de 100k descriptores locales, obtenga el **codebook** aplicando clustering K-Means (se recomienda utilizar la implementación de OpenCV). Deben generar tres **codebooks** de diferentes tamaños: 64, 128 y 256 clusters. Recuerden que VLAD requiere un **codebook** de menor tamaño que el requerido por BoF.

3) Agregar los descriptores usando VLAD según [2]. Con esto, cada imagen queda representada por un descriptor VLAD, el cual es de tamaño KxD, donde K=64, 128 o 256 y D =128 (por SIFT).

4) Con los descriptores VLAD calculados, ordene el *dataset* con respecto a la distancia calculada sobre la consulta, de menor a mayor, lo cual producirá el ranking de recuperación final. Como función de distancia utilice la distancia Euclidiana y Hellinger.

5) Por cada consulta, calcule la precisión promedio (average precision - AP). Usando la AP por cada consulta, calcule la precisión media con respecto a todas las consultas (Mean Average Precision - mAP) y genere el gráfico Precision-Recall. Presente y analice resultados para K=64, 128 y 256 usando distancia Euclidiana y Hellinger.

6) **Entrega**
   a) Entregar Informe tipo *paper* incluyendo: Introducción, Diseño y Desarrollo, Evaluación Experimental y Análisis de Resultados, y Conclusiones.
   b) Como de costumbre, la tarea se desarrolla en equipos de 2 estudiantes.
   c) Fecha de entrega: hasta **24 de mayo, 23:59**. No se aceptan atrasos (se considera 1 hora como tiempo de gracia) . Enviar la tarea al ayudante Erick Svec <evsvec@uc.cl>.

7) **Referencias**

[1] *Philbin, J. , Chum, O. , Isard, M. , Sivic, J. and Zisserman, A. "***Object retrieval with large vocabularies and fast spatial matching".** Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[2] Jegou, H.; Perronnin, F.; Douze, M.; Sanchez, J.; Perez, P.; Schmid, C., "**Aggregating Local Image Descriptors into Compact Codes**", *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.34, no.9, pp.1704,1716, Sept. 2012.