

# Lab 7: PHOW Classification

Duván Alberto Gómez Betancur  
Visión Artificial  
Universidad de los Andes  
da.gomez16@uniandes.edu.co

## Abstract

*Este informe presenta el resumen de los resultados obtenidos en el laboratorio 7 del curso de Visión Artificial, cuya temática específica es la clasificación PHOW. El documento comprende la ejecución de diferentes pruebas del algoritmo sobre las bases de datos Imagenet y Caltech101 variando diferentes parámetros. Se presentan la parametrización, la metodología de las pruebas realizadas y los resultados obtenidos, considerando los tiempos de cómputo y evaluándolos a nivel de la precisión representada en una matriz de confusiones para cada caso probado.*

## 1. Introducción

La clasificación de imágenes por “categorías” a partir de los objetos que contienen, es decir, la clasificación de imágenes por categorías de objetos, ha sido un problema ampliamente estudiado y para el cual se han propuesto diversidad de soluciones e incluso diversidad de bases de datos de entrenamiento y prueba, tratando cada vez más de operar sobre imágenes más naturales y con un mayor número de categorías.

Una de las soluciones propuestas al problema de la clasificación de imágenes por categorías de objetos es mediante la representación Pyramid Histogram Of visual Words (PHOW) que es una variante de los descriptores “densos” de Scale-invariant feature transform (Dense SIFT), extraídos a múltiples escalas.

Los representaciones PHOW combina descriptores de apariencia y forma junto con información espacial dentro de la imagen.

En este laboratorio el uso del algoritmo de clasificación a través de la representación PHOW se realizó utilizando el código del archivo denominado *phow\_caltech101.m*, escrito por Andrea Vedaldi y que hace parte de la librería VLFeat.

La librería VLFeat implementa algoritmos populares de visión por computador especialmente en análisis de imágenes, extracción de características locales en imágenes y *emparejamiento*. Se encuentra disponible en

<http://www.vlfeat.org/index.html> y es un desarrollo de tipo open source.

Las bases de datos utilizadas en este laboratorio para probar el funcionamiento de *phow\_caltech101.m* fueron Caltech 101 y Imagenet.

Este informe está organizado de la siguiente manera: en la sección II se describen los materiales y métodos utilizados, en la sección III se presentan los resultados obtenidos tanto sobre la base de Caltech101 como sobre la base de Imagenet, y finalmente en la sección IV se presenta la discusión de los resultados obtenidos y las posibles mejoras.

## 2. Materiales y métodos

### 2.1. Base de datos

Para el desarrollo del laboratorio se utilizaron dos bases de datos, a saber: Imagenet y Caltech 101.

La base de datos Caltech 101 está compuesta por imágenes de objetos pertenecientes a 102 categorías. Caltech 101 tiene aproximadamente 50 imágenes de tamaño 300 x 200 por categoría. La organización de la base es por carpetas, cada una pertenece a una categoría y contiene las imágenes respectivas.

En la versión de la base de datos Imagenet utilizada se tiene un conjunto de imágenes *naturales* divididas en 200 categorías con aproximadamente 100 imágenes cada una de ellas. Esta base al igual que Caltech 101 está organizada como un conjunto de carpetas donde cada una representa una categoría.

Es importante mencionar que tanto Caltech 101 como Imagenet están organizadas por categorías de objetos con una carpeta por categoría como lo requiere el código de *phow\_caltech101.m*.

### 2.2. Método

Como se mencionó anteriormente el desarrollo del presente laboratorio se basó específicamente en la implementación del código *phow\_caltech101.m* que hace parte de la librería de código abierto VLFeat y desarrollado por Andrea Vedaldi.

El código de Vedaldi de manera análoga al algoritmo de clasificación por textones visto en la práctica de laboratorio anterior, está compuesto por pasos que van desde la extracción de descriptores, construcción de representación o diccionario de palabras, construcción de la representación de cada imagen y posterior clasificación.

Específicamente el código comprende: i) Extracción de descriptores Dense SIFT de las imágenes, ii) Construcción de la representación PHOW y agrupamiento para formar un diccionario de palabras visuales (600 para este caso), iii) A partir del diccionario de palabras visuales se realiza una asociación con los descriptores y calcula a manera de histograma la representación de cada imagen, iv) Finalmente se entrena un clasificador, en este caso se utilizaron máquinas de soporte vectorial (SVM por sus siglas en inglés).

Todos los pasos están implementados de manera congruente en el código proporcionado por Vedaldi, que se encuentra parametrizado para operar satisfactoriamente sobre la base de datos Caltech 101.

En el caso de Caltech 101 se corrieron pruebas en la versión completa y en la versión tiny del código y de la base de datos.

Por lo anterior fue necesario modificar el código de *phow\_caltech101.m* a fin de poderlo operar sobre las imágenes de la base de datos Imagenet

Los cambios realizados sobre el código fueron: 1) Cambio en la variable que define el directorio donde se encuentran las imágenes, 2) Cambio en la variable que define el directorio donde se guardan los resultados y 3) Cambio en la definición de la extensión de las imágenes a probar, definiéndola como .JPEG.

### 3. Pruebas y Resultados

Esta sección detalla las pruebas y los resultados de las mismas. Se evaluaron los resultados a nivel de precisión a partir de la matriz de confusiones para cada caso.

Específicamente se probó el comportamiento del algoritmo al modificar las variables: número de imágenes de entrenamiento *conf.numTrain*, número de clases *conf.numClasses* y el número de particiones espaciales *conf.numSpatialX*.

A continuación se presentan los resultados obtenidos para cada base de datos con los diferentes parámetros evaluados. Vale la pena mencionar que las pruebas se corrieron en un servidor con Fedora release 20, 62 GB de memoria RAM y Matlab 2012.

Los resultados se presentan en Figuras que comprenden dos gráficas, una asociada al score asignado por el algoritmo a las clases y otra perteneciente a la matriz de confusiones.

#### 3.1. Pruebas sobre Caltech 101

Como se mencionó anteriormente la pruebas corrientes sobre cada base de datos comprendieron la variación

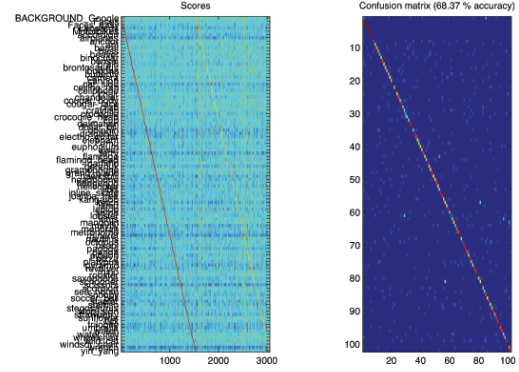


Figure 1. Resultados parámetros baseline sobre Caltech 101 completa

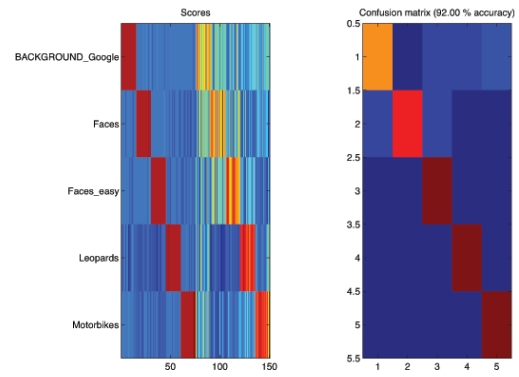


Figure 2. Resultados parámetros baseline sobre Caltech 101 Tiny

de los parámetros *conf.numTrain*, *conf.numClasses* y *conf.numSpatialX*

Los resultados se evaluaron utilizando la matriz de confusiones que provee el código desarrollado por Andrea Vedaldi.

A continuación se presentan los resultados obtenidos:

En la Figura 1 se presentan los resultados sobre la base completa de Caltech 101 con los parámetros de base presentados en el código y en la Figura 2 se presentan los resultados sobre la versión Tiny de la base con los mismos parámetros.

En las Figuras 3 a 8 se presentan los resultados variando los parámetros *conf.numTrain*, *conf.numClasses* y *conf.numSpatialX*, con dos pruebas para cada caso respectivamente.

#### 3.2. Resultados sobre Imagenet

En el caso de Imagenet se realizó una variación de parámetros similar a la de Caltech 101. Se variaron los parámetros *conf.numTrain*, *conf.numClasses* y *conf.numSpatialX*

En este caso nuevamente los resultados se evaluaron utilizando la matriz de confusiones que provee el código de-

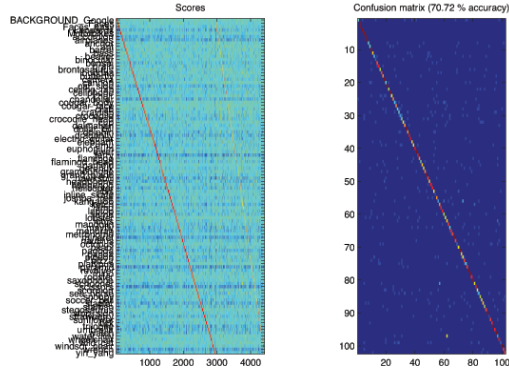


Figure 3. Resultados variación parámetro `conf.numTrain` a 30 sobre Caltech 101

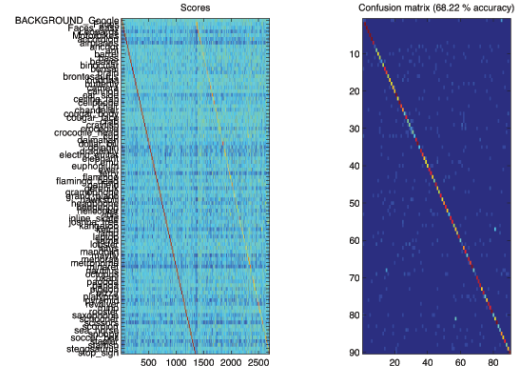


Figure 6. Resultados variación parámetro `conf.numClasses` a 90 sobre Caltech 101

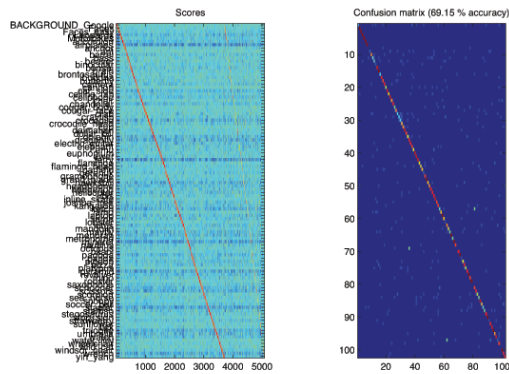


Figure 4. Resultados variación parámetro `conf.numTrain` a 40 sobre Caltech 101

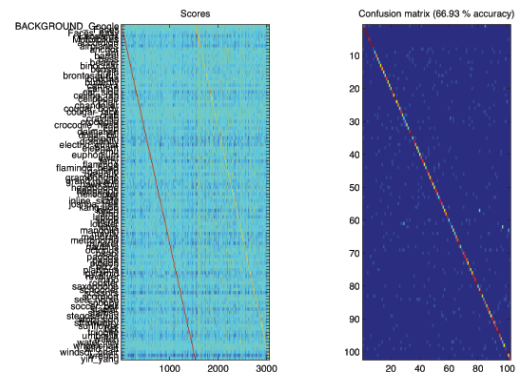


Figure 7. Resultados variación parámetro `conf.numSpatialX` a [4 6] sobre Caltech 101

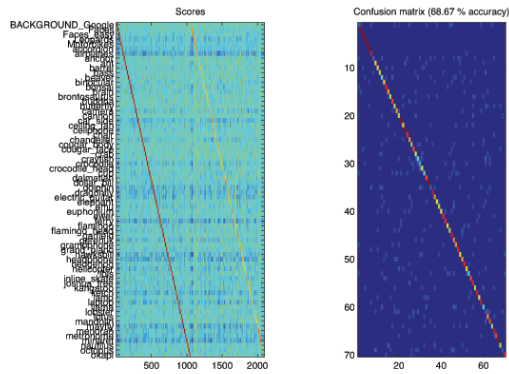


Figure 5. Resultados variación parámetro `conf.numClasses` a 70 sobre Caltech 101

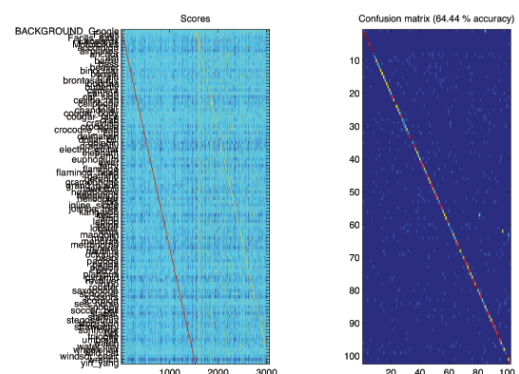


Figure 8. Resultados variación parámetro `conf.numSpatialX` a [6 8] sobre Caltech 101

sarrollado por Andrea Vedaldi.

A continuación se presentan los resultados obtenidos:

En la Figura 9 se presentan los resultados sobre la base completa de Imagenet con los parámetros de base ajustados.

En las Figuras 10 a 16 se presentan los resultados variando los parámetros `conf.numTrain`, `conf.numClasses` y `conf.numSpatialX`.

## 4. Discusión y posibles mejoras

Los resultados obtenidos en la práctica de laboratorio fueron evaluados a partir del valor de precisión promedio (valores de la matriz de confusiones) entregado por el código de Vedaldi.

Vale la pena mencionar que los valores seleccionados

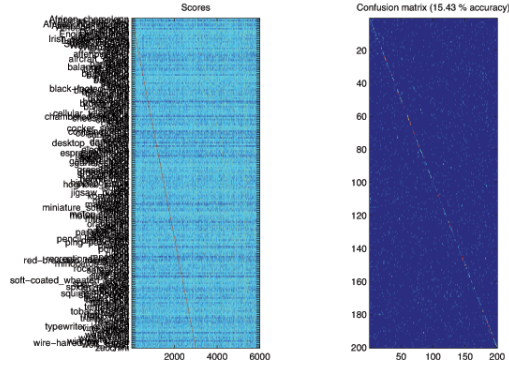


Figure 9. Resultados parámetros baseline sobre Imagenet

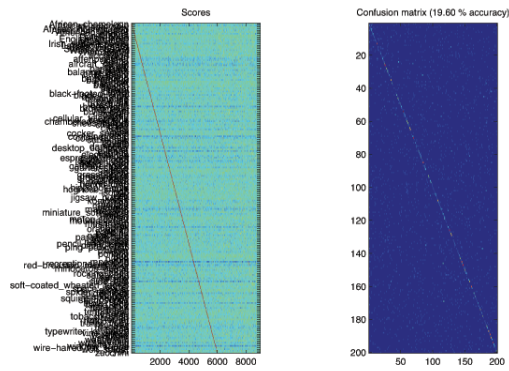


Figure 10. Resultados variación parámetro *conf.numTrain* a 30 sobre Imagenet

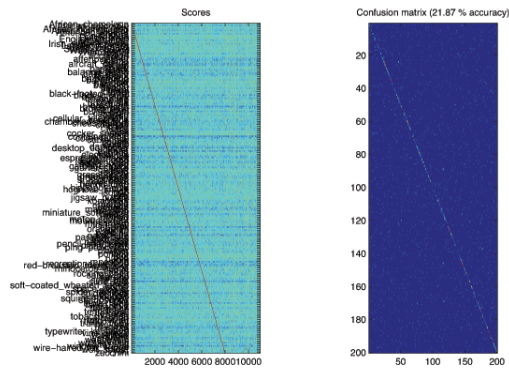


Figure 11. Resultados variación parámetro *conf.numTrain* a 40 sobre Imagenet

para las diferentes variables en las pruebas implementadas, fueron elegidos de manera arbitraria con el ánimo de dar respuesta a los interrogantes presentados en la guía del laboratorio en cuanto a la variación específica de dichas variables.

Para el caso de Caltech 101 como se puede observar en los resultados presentados en la sección anterior para la base de datos Caltech 101 se logró una precisión promedio de

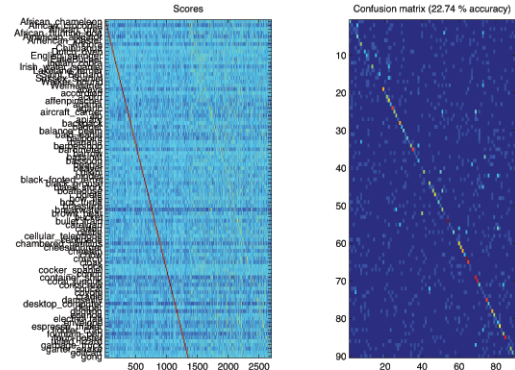


Figure 12. Resultados variación parámetro *conf.numClasses* a 90 sobre Imagenet

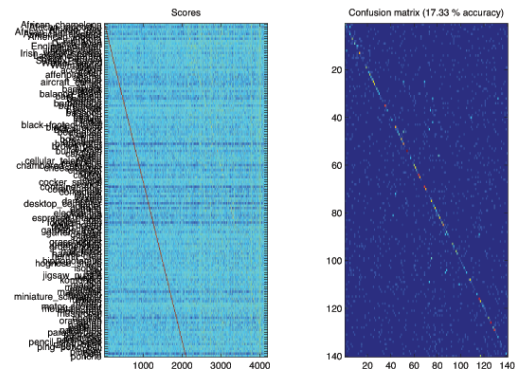


Figure 13. Resultados variación parámetro *conf.numClasses* a 140 sobre Imagenet

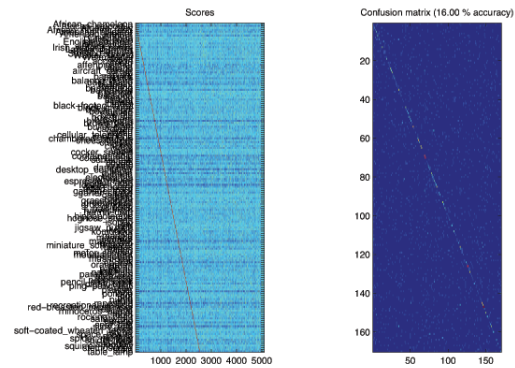


Figure 14. Resultados variación parámetro *conf.numClasses* a 170 sobre Imagenet

92% con la versión Tiny de la base y una precisión promedio de 68.37% con la base de datos completa.

Ahora bien en la base Caltech 101 completa variando el parámetro *conf.numTrain* a 30 y 40, se obtuvieron resultados de precisión promedio de 70.72% y 69.15% respectivamente.

En esta misma base en cuanto a los cambios en los



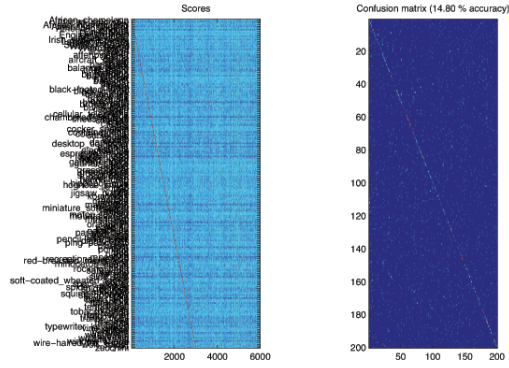


Figure 15. Resultados variación parámetro `conf.numSpatialX` a [4 6] sobre ImageNet

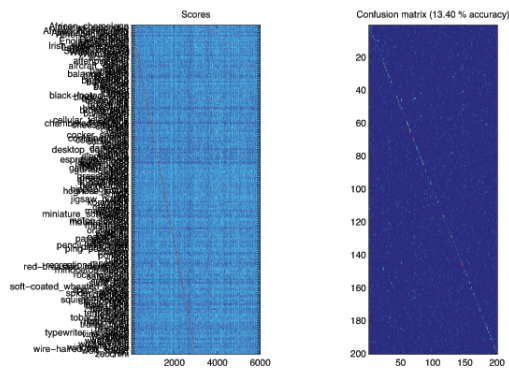


Figure 16. Resultados variación parámetro `conf.numSpatialX` a [6 8] sobre ImageNet

Parámetro	Precisión promedio (%)
<code>conf.numTrain=30</code>	70.72
<code>conf.numTrain=40</code>	69.15
<code>conf.numClasses=70</code>	68.67
<code>conf.numClasses=90</code>	68.22
<code>conf.numSpatialX=[4 6]</code>	66.93
<code>conf.numSpatialX=[6 8]</code>	64.44

Table 1. Precisión Nearest Neighbor para  $k=10, 20$  y  $40$

parámetros seleccionados para variación se tienen los valores presentados en la Tabla 1

A partir de las variaciones en los valores de precisión relacionados en la Tabla 1 y considerando los valores obtenidos con los parámetros que esta por defecto en el código, es posible afirmar que el método de clasificación por representación PHOW es ciertamente sensible a los parámetros utilizados.

Para el caso de Caltech 101 la mejor precisión promedio se obtuvo al definir un conjunto de entrenamiento compuesto por 30 imágenes y dejando invariantes los demás parámetros.

Asimismo, es claro que el parámetro que generó los resultados más bajos en precisión promedio fue `conf.numSpatialX` lo que podría demostrar una alta dependencia del algoritmo a este parámetro, específicamente en el momento de la extracción de los descriptores.

Para ImageNet se corrieron pruebas similares a las descritas para Caltech 101, logrando la mayor precisión al variar únicamente el parámetro `conf.numClasses` y definirlo en 90.

Nuevamente para ImageNet, el peor valor de precisión promedio se dio al variar el parámetro `conf.numSpatialX`, reforzando un poco la hipótesis de dependencia alta del algoritmo a este parámetro, específicamente en el momento de la extracción de los descriptores.

La limitación en el número de pruebas realizadas se debió especialmente a que cada una requirió bastante tiempo de procesamiento y capacidad de cómputo debido al alto número de imágenes a analizar. Por esta razón se limitó el número de pruebas realizadas.

Finalmente, posibles mejoras a los resultados obtenidos en el presente laboratorio podrían venir, de manera análoga al laboratorio anterior, de incrementar el número de observaciones (imágenes) por clase para el entrenamiento y limitar un poco el número de categorías.

Además podría introducirse información de color o textura u otro tipo de descriptores que permita una mejor clasificación al entregar más información de cada imagen. Asimismo es posible analizar el paso del agrupamiento para la creación del diccionario de palabras visuales y tal vez definir uno mejor, al igual que aunque no se probó en la presente práctica, probar diferentes tamaños del diccionario de palabras visuales para a fin de tener una mejor clasificación de las imágenes de prueba.

## References

- [1] Bosch, A.; Zisserman, A.; Muñoz, X., "Image Classification using Random Forests and Ferns," Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on , vol., no., pp.1,8, 14-21 Oct. 2007