

Laboratory 8: PHOW

Fabian Martínez
Universidad de los Andes
Bogota

fe.martinez10@uniandes.edu.co

Juan Felipe Pérez
Universidad de los Andes
Bogota

jf.perez10@uniandes.edu.co

Abstract

En el siguiente informe se presentan los resultados obtenidos al realizar la clasificación de imágenes por medio de PHOW. Inicialmente se detalla el de algoritmo y de los parámetros mas relevantes involucrados en el procedimiento, luego se presentan unos resultados obtenidos con la base de datos de Caltech y por último los resultados de la base ImageNet200 junto con un pequeño análisis.

1. Introducción

La clasificación de las imágenes es un proceso que es utilizado en una gran cantidad de áreas del conocimiento como por ejemplo en la cartografía, en la medicina al realizar diagnósticos, en la astronomía, etc. En general, el proceso de clasificación se basa en la capacidad de extraer ciertas características de las imágenes, con el fin de posteriormente clasificarlas en las diferentes clases posibles.

Actualmente existe una gran cantidad de métodos y de procedimientos diferentes que utilizan diferentes características de las imágenes, cada uno de los cuales tiene un alcance diferente. Uno de los principales métodos corresponde a PHOW, el cual es un modelo que representa las características de la imagen como una bolsa de palabras, y calcula la frecuencia en que se repiten estas características en la imagen. Para solucionar el problema de información espacial, se divide la imagen en subregiones, con lo cual es posible formar una pirámide.

Con el fin de evaluar la eficiencia de este método y de los diversos parámetros, en el siguiente laboratorio se realiza una evaluación en dos bases de datos diferentes, la de Caltech y la ImageNet200.

1.1. Objetivos

- Entender el funcionamiento de la clasificación de imágenes por medio de PHOW.
- Obtener los parámetros de PHOW óptimos para realizar la mejor clasificación de las imágenes en las

diferentes clases.

- Realizar un entrenamiento en las bases de Caltech y de ImageNet200, de manera que se puedan obtener comparaciones entre los resultados obtenidos.

2. Materiales y Métodos

Para el siguiente laboratorio se utilizaron códigos y funciones en Matlab junto con las bases de datos de Caltech y de ImageNet.

2.1. Bases de datos

Para el laboratorio se utilizaron dos bases de datos, la primera de ellas corresponde a la de Caltech. Esta base de datos está compuesta por 102 categorías de imágenes, en las cuales la cantidad de imágenes en cada categoría es diferente, como por ejemplo en la de aviones se encuentran 800 imágenes, mientras que en la de cocodrilos hay solo 50 imágenes. Por otra parte, el tamaño de las imágenes puede ser diferente, y se encuentran en formato .jpg. La segunda base de datos corresponde a ImageNet. En esta base de datos se tiene una mayor cantidad de clases de imágenes, de las cuales se van a utilizar 200, y se tiene una mayor cantidad de imágenes por cada clase. Cabe resaltar que en esta base de datos las imágenes se encuentran separadas en un conjunto de Train y en un conjunto de Test, y el formato es .jpeg.

2.2. Método PHOW

Hace parte de la librería que tiene los códigos de las funciones utilizadas en la visión por computador. El método de PHOW tiene las siguientes secciones: Inicialmente se tiene la configuración de los parámetros mas importantes como por ejemplo la cantidad de palabras o de imágenes de entrenamiento, luego se descarga la base de datos de Caltech en caso de que esta no exista. A continuación se realiza la configuración teniendo en cuenta la base de datos (nombre de las imagenes, clases, etc). Posteriormente se realiza el entrenamiento de la bolsa de palabras y se realiza el vector

de cuantización de las mismas. Luego se realiza una computación teniendo en cuenta la información espacial para finalmente realizar el entrenamiento de los Support Vector Machine para la clasificación de las imágenes. Una vez se tienen los SVM entrenados, se realiza la clasificación de las imágenes de test y se crea la matriz de confusión donde se evalúa la eficiencia del método de clasificación.

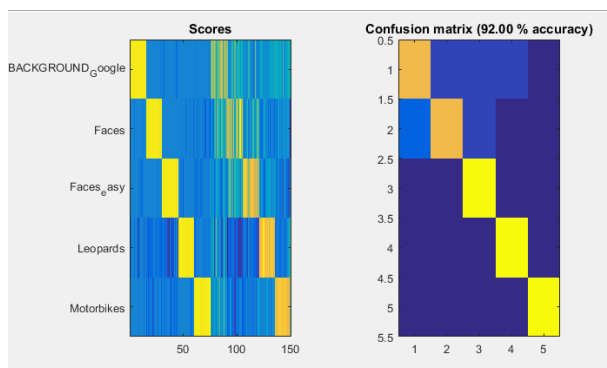
2.3. Parámetros de cambio

En el algoritmo existen varios parámetros, entre los principales se encuentran: la cantidad de imágenes que se utilizan en entrenamiento y en prueba, la cantidad de clases que se toman para la ejecución del algoritmo y la posterior construcción de la matriz de confusión, el número de palabras o descriptores de la imagen, la configuración espacial para formar las pirámides, el valor de la variable de holgura para entrenar los SMV. Además existen otros parámetros dentro del código, como por ejemplo al realizar el mapa de características, o el método de entrenamiento de los SMV.

3. Resultados

3.1. Base de datos de Caltech

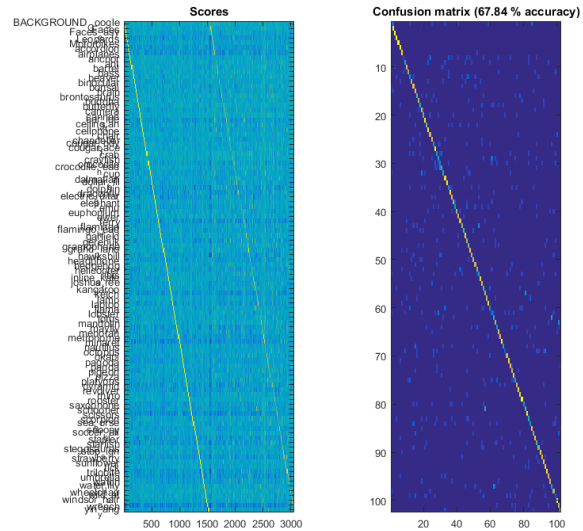
Como primera medida, se corre el código de Matlab en una versión tiny, es decir, que no se toma en cuenta de manera completa la base de datos de Caltech, sino únicamente unas pocas clases de imágenes (5 clases). Los resultados que se obtuvieron se presentan en la figura 3.1



Como se puede apreciar en la imagen anterior, la precisión de este método en este problema de pocas clases es de aproximadamente el 92%. Nuevamente se ejecuta el código, pero ahora con todas las diferentes clases (102), y se tienen los siguientes resultados.

En la imagen ?? es posible apreciar que la precisión del método bajó considerablemente, llegando a ser de solo el 67%, dejando los demás parámetros por default. A continuación se evalúa el efecto de los diferentes parámetros en los resultados de clasificación.

- Número de imágenes de entrenamiento Se varía el número de imágenes de entrenamiento, iniciando con



un número pequeño y aumentándolas gradualmente. A medida que se aumenta la cantidad de entrenamiento, se tiene una mejor precisión, sin embargo puede llegarse a un punto en el que hay overfitting en la clasificación y empieza a tenerse nuevamente un alto error.

- Número de clases Es un parámetro que influye bastante, puesto que una mayor cantidad de clases indica una mayor dificultad en la clasificación, por lo tanto necesita una buena cantidad de descriptores para poder diferenciar las clases. Por otro lado, aumentar la cantidad de clases es beneficioso ya que reduce la posibilidad de clasificar correctamente por azar, sin embargo aumenta la cantidad computacional.
- Número de palabras En general, se tiene el siguiente comportamiento: aumentar la cantidad de palabras se tiene una mejor estimación de la clase a la que pertenecen las imágenes, sin embargo llega a un punto de saturación en el que aumentar el número de palabras no es viable o no es muy eficiente.
- Variación de C A pesar de que este parámetro se cambió considerablemente (desde 0.1 hasta 100) la precisión obtenida con el método de PHOW no se ve afectada, puesto que la variación en el resultado entre estos valores no fue mayor al 2%.
- Posición X – Y Es de relevancia para la construcción de las pirámides que se utilizan en el algoritmo. Las variaciones en la precisión son diversa, y no se logra llegar a un único punto óptimo.

Finalmente, los parámetros a elegir para el entrenamiento en ImageNet se decidieron en base a los resultados obtenidos

anteriormente, es decir, en aquellos que mostraban un mejor rendimiento y precisión en la clasificación. Así mismo se tuvo en cuenta la cantidad de imágenes tanto en entrenamiento como prueba, y que variables como por ejemplo la cantidad de palabras no fueran exageradamente grandes para no tener grandes tiempos de computación.

3.2. ImageNet200

A continuación se pueden observar las matrices de confusión resultantes de los procesos de train y test en imageNet200.

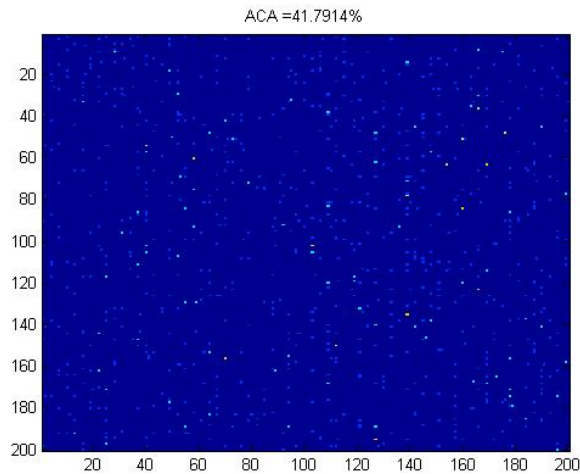


Fig. 1. Matriz de confusión del resultado en train, con su respectivo ACA

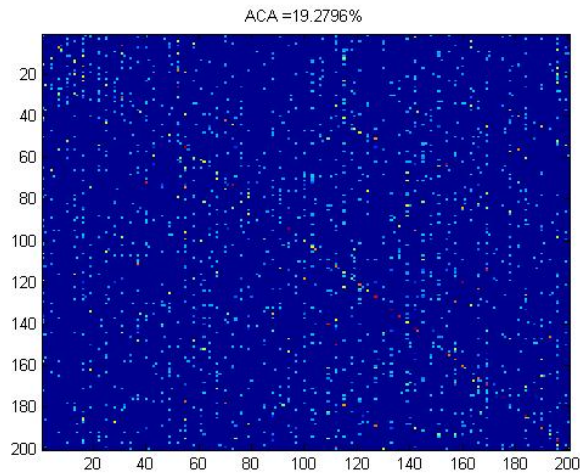


Fig. 2. Matriz de confusión del resultado en test, con su respectivo ACA

Como se puede observar, los resultados en imageNet son claramente inferiores a los obtenidos en caltech101. Esto

puede ser debido a que el tamaño de la base de datos de imageNet es mucho más grande, y este tiene el doble de categorías de caltech. Esto hace que el modelo sea más propenso al error, pues habiendo más categorías, más difícil elegir la categoría correcta. Otro factor que pudo influir bastante es que para imagenet, debido a su gran extensión, se realizó un subsample a las imágenes. De esta forma pudo haber sido omitida información relevante para el modelo.

4. Conclusiones

- Como era de esperarse, el cambio en algunos parámetros afecta de manera significativa la efectividad del método.
- La efectividad de un método depende no solo del método como tal, sino que, como vimos en laboratorio, un mismo método puede cambiar radicalmente de efectividad con solo cambiar de set en donde se pruebe.

5. Bibliografía

1. Mahdi, S. What you need to know about the state-of-art computational models of object-vision. Universidad de Cambridge.
2. Yang, Y y Newsam, S. Bag of visual words and spatial extensions for land-use classification.