

Clasificación de imágenes de la base de datos imagenet mediante el metodo de PHOW

Cristian Alejandro Vergara Perico
Universidad de los Andes, Bogotá, Colombia.

Keywords—*Clasificación, Imagenet, SVM, PHOW,*

I. DESCRIPCIÓN DE LA BASE DE DATOS

LA base de datos contiene 20000 imágenes descargadas de internet, lo que garantiza la generalización de la base de datos, dado que son imágenes tomadas por cualquier individuo. Estas imágenes son clasificadas en 200 categorías diferentes.

Para este informe las imágenes son empleadas para el entrenamiento y la evaluación del método.

II. DESCRIPCIÓN Y RECONOCIMIENTO DEL MÉTODO

El método fue desarrollado por Andrea Vedaldi y se encuentra documentado en [1]. Se basa en la clasificación usando como características una variación del método SIFT. Esta es llamada PHOW, la cual se basa en el mismo método SIFT con variaciones que logran reducir el costo y el tiempo computacional. Estos métodos se basan en encontrar descriptores invariantes a la escala y a la rotación, mediante la búsqueda de blobs en imágenes. Estos puntos reúnen información espacial y la información de la orientación del gradiente normalizada de los pixeles que se encuentran alrededor, con el fin de obtener una información mas relevante a este keypoint esta información es normalizada mediante una curva gaussiana.

La información de orientación es agrupada en 600 grupos mediante el método de k-means, con lo cual se encuentra el vocabulario de las orientaciones. A partir de este vocabulario se realiza un histograma de cómo se distribuyen los descriptores de la imagen en las 600 palabras.

El método además añade información espacial de los descriptores mediante la formación de histogramas que contienen la ocurrencia en diferentes sectores de la imagen, el número de sectores es variable por el usuario.

Lo primero que realiza el método es la extracción de la información que se va a clasificar, de esta manera se deben establecer el número de categorías, la cantidad de imágenes de entrenamiento, la cantidad de imágenes de evaluación y el número de particiones de la imagen para encontrar la información espacial.

A partir de esto el algoritmo selecciona imágenes aleatoriamente perteneciente a un numero de categorías establecido, teniendo encuentra el número de imágenes de entrenamiento y de test. A estas imágenes se les aplica el método de PHOW con el fin de extraer los histogramas representativos de cada imagen.

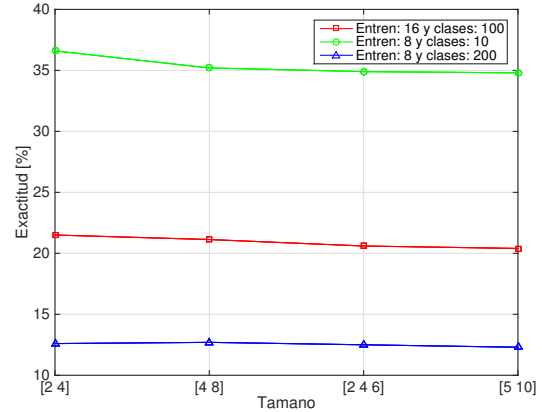


Figura 1: Sencibilidad del metodo al cambio del número de particiones de la imagen.

Una vez se obtiene el histograma de cada imagen se concatenan los histogramas de todas las imágenes tanto de entrenamiento como de evaluación. De esta manera es posible formar el kernel, el cual se usará en método de clasificación support vector machines (SVM).

En la etapa de clasificación primero se debe generar el modelo de SVM, para este proceso se emplean los histogramas extraídos de las imágenes de entre entrenamiento. Una vez obtenido del modelo se emplea para la clasificación de las imágenes de test.

III. RESULTADOS DE ENTRENAMIENTO Y TEST

Primero se encontro la sencibilidad del cambio de tamaño los resultados se encuentran en la figura 1

Para encontrar la sencibilidad del metodo a variaciones del número de imagenes de entrenamiento y número de clase, se realizaron 16 mediciones en las cuales se encontró la exactitud para cuatro numero de imágenes de entrenamiento: [2 8 16 25] para cada uno se vario el número de clases, [10 50 100 200]. Los resultados se encuentran en la figura 2.

Los tiempos de procesamiento se ecuentran resumidos en la tabla xxxx

IV. DISCUSIÓN DE RESULTADOS

IV-A. Bases de datos

De encontró que los resultados obtenidos en la base de datos de Caltech 101 mostraron un mejor desempeño que los

¹Este trabajo fue realizado para el curso de maestría Visión Artificial dictado el primer semestre de 2015 en la Universidad de los Andes, Colombia

		Clases			
		10	50	100	200
Entrenamiento	2	10.5	17.4	24.3	27.3
	8	12.3	19.6	26.8	29.8
	16	14.5	20.6	28.9	33.5
	25	18.2	22.3	30.1	38.3

Cuadro I: Tiempo en minutos de procesamiento. Número de particiones espaciales [2 4].

obtenidos en la base de datos IMAGINET TINY. En la figura 2 se puede observar que la exactitud de los resultados obtenidos en Caltech 101 corresponden aproximadamente a 1.5 de los resultados obtenidos en imagenet.

Este fenómeno se puede atribuir al hecho de que la base de datos imagenet contiene imágenes mas aleatorias, es decir, son imágenes tomadas por cualquier usuario, mientras que en el caso de Caltech 101 son imágenes mas específicas en captar objetos pertenecientes a cada una de las categorías .

IV-B. Variación de parámetros

En la figura 1 se observa que la variación del tamaño del número de particiones de la imagen no tubo una variación significativa del desempeño del clasificador, al aumentar los grupos tubo un pequeño declive en el desempeño en forma lineal , además el tiempo y costo computacional también incrementa drásticamente, de esta manera se decidió dejar constante este parámetro durante la variación del número de imágenes de entrenamiento y de las clases a evaluar.

Se encontró que al incrementar el número de imágenes de entrenamiento el desempeño incrementaba, sin embargo este aumento tiende a decrecer al aumentar el número de imágenes de entrenamiento como se evidencia en la figura 2. Esto se puede atribuir a que con un mayor número de imágenes de entrenamiento, el algoritmo tiene un mayor aprendizaje, logrando tener en cuenta mayor numero de discriminadores .

Por otro lado al aumentar el número de clases la exactitud disminuyo de mera exponencial. Esto se dio gracias a que

existe un mayor número de posibilidades de clasificar mal una imagen con lo que la probabilidad de tener éxito disminuye.

IV-C. Recursos computacionales

En el cuadro I se observa que conforme aumenta el número de imagenes de entrenamiento y de clases el tiempo empleado es mucho mayor, de igual manera al procesar el algoritmo en un computador ordinario con 200 categorías y 25 imagenes de entrenamiento la memoria del computador fue insuficiente, de tal manera se decidió emplear un servidor (Guitaca).

V. LIMITACIONES DEL MÉTODO

El costo computacional y el tiempo de procesamiento del método es bastante alto, por lo que se requiere un computador de gran capacidad cuando se requieren procesar bases de datos de gran tamaño.

Aunque 200 es un gran número de categorías se desearía probar el método en imágenes mas aleatorias.

VI. MEJORAS DEL METODO

Este método es una variación simplificada del método SIFT para aumentar la velocidad de procesamiento mientras se disminuye el costo computacional. Por lo que para mejorar el desempeño se recomienda usar el método SIFT.

Se aconseja aumentar la base de datos se puede generalizar mas el entrenamiento y mejorar la clasificación de las imágenes.

Aunque el modelo de SVM es bastante preciso a la hora de clasificar, es un metodo que tiene un alto costo computacional [2]. Para disminuir el costo computacional, se recomienda probar con metodos como arboles de desición o nearest neighbour a expensas de los resultados que se puedan obtener.

REFERENCIAS

- [1] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," <http://www.vlfeat.org/>, 2008.
- [2] MATLAB, version 8.4.0 (R2014b). Natick, Massachusetts: The Math-Works Inc., 2014.

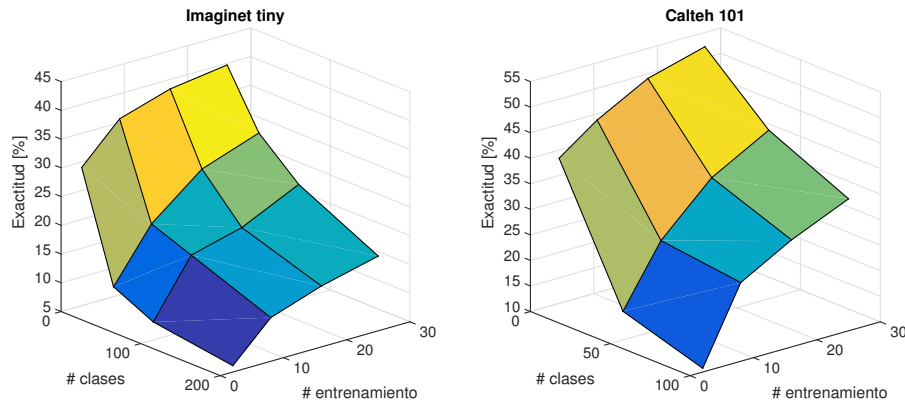


Figura 2: Sensibilidad del metodo al cambios del número de clases empleadas y el numero de muestras de entrenamimento