

Laboratory 9: HOG

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 presenta inicialmente una breve descripción de el método HOG (histograma de gradientes orientados) el cual es un descriptor altamente usado en visión por computador con el objetivo de detectar objetos. Posteriormente se presentan los resultados obtenidos de utilizar este proceso en la identificación de caras en la base de datos WiderFace. Por último se realizan algunos análisis y comentarios sobre los resultados obtenidos.

1. Introducción

La detección de objetos es un problema en el mundo de la visión por computador en el cual se busca encontrar y/o detectar un objeto en una imagen o incluso en una secuencia de video.

Uno de los casos mas específicos y mas estudiados en el área de detección corresponde a la identificación de rostros humanos en una imagen, para lo cual se han realizado una gran cantidad de algoritmos y métodos que a pesar de que se han logrado unos resultados bastante aceptables, el proceso no es tan sencillo a como lo realiza el sistema visual humano.

En los problemas de detección pueden presentarse una gran cantidad de problemas en las imágenes como lo son el punto de vista del observador, la iluminación, la oclusión y la escala que pueden presentar las diferentes caras de las personas que se encuentran en la imagen.

Para el siguiente laboratorio se utilizará el método de HOG a distintas escalas de las imágenes con el objetivo de obtener una descripción de las imágenes para entrenar un clasificador basado en una máquina de soporte vectorial(SVM) junto con un proceso de "hard negative mining" para minimizar las detecciones falsas con el objetivo de identificar caras.

1.1. Objetivos

- Entender y utilizar el método de HOG para la detección de objetos.

- Detectar las caras de las personas en diferentes tipos de imágenes.
- Entrenar un clasificador basado en HOG para la clasificación de las detecciones.

2. Materiales y Métodos

En el presente laboratorio se utilizó como herramienta computacional Matlab, además de las imágenes que fueron obtenidas de la base de datos WiderFace.

2.1. Base de datos

La base de datos utilizada para el proyecto es WIDER FACE: A Face Detection Benchmark. Esta contiene imágenes de caras que tienen las etiquetas correspondientes. Una de las características principales de esta base de datos es que se garantiza una alta variabilidad entre las imágenes, con cambios en las escalas, poses, oclusiones y ambientes en los que se encuentran los objetivos. La base de datos a utilizar se encuentra dividida en tres secciones, una sección de imágenes de entrenamiento con solo la cara de las personas, una sección de entrenamiento en la que se encuentran personas en los respectivos ambientes y finalmente una sección para el test. Cada una de estas secciones se encuentra compuesta por 62 clases diferentes.

2.2. HOG

El histograma de gradientes orientados (o HOG por sus siglas en ingles) es un descriptor que es utilizado altamente en la visión por computador con el objetivo de detectar objetos dentro de una imagen. La idea principal de esta técnica radica en que la forma de un objeto específico puede ser descrita a través de la distribución de los gradientes de borde.

Los pasos que se realizan para la implementación para la detección y clasificación de peatones utilizando HOG son los siguientes:

- La carga de los datos (imágenes) de la base de datos, tanto de los ejemplos positivos como de los ejemplos negativos.

- Se configura la extracción de los parámetros HOG junto con el hiperparámetro del número de escalas a utilizar.
- Extraer las características de HOG a las imágenes de entrenamiento de la base de datos en las diferentes escalas.
- Se obtienen los candidatos a los cuales se les va a realizar la clasificación, entre los cuales pueden encontrarse varios objetivos dentro de una misma imagen. En estos casos se utiliza un algoritmo de supresión de no máximos para evitar sobrelapar las posibles diferentes detecciones de una imagen.
- Se realiza un entrenamiento de un clasificador. En este caso el clasificador es un Support Vector Machine, el cual es óptimo para realizar clasificaciones binarias (es o no es una cara).
- Con el fin de tener un mejor desempeño en la clasificación con SVM, se modifica el modelo obtenido por medio de Hard negative mining, es decir, se obliga a que los errores obtenidos se utilizar para reentrenar el modelo iterativamente hasta que estos queden clasificados correctamente.
- Finalmente se evalúa el modelo obtenido en las imágenes de test de las base de datos y se obtiene un resultado de su desempeño.

2.3. Hiperparámetros

Uno de los parámetros que se puede modificar en el proceso de detección de caras corresponde al número de tamaños en los cuales se evalúa la imagen al utilizar HOG, es decir, la cantidad de niveles en la pirámide de la imagen.

La utilidad que presenta este parámetro es la posibilidad de detectar caras de personas que se encuentren en diferentes tamaños (escalas) en una misma imagen como en diferentes imágenes.

2.4. Evaluación

La evaluación de la detección de imágenes de caras se realiza con una curva de AP (curva de precisión / cobertura). En esta se evalúan el número de detecciones que se realizaron correctamente sobre el número total de detecciones debieron ser realizadas, junto con la cantidad de detecciones se realizaron correctamente sobre la cantidad de detecciones que debían obtenerse.

Una vez evaluadas las detecciones, es posible evaluar que tan correctas se realizaron las detecciones (tamaño de los recuadros), para lo cual se utiliza el índice de Jaccard, es decir, el índice que indica que tanto se sobrelapa la zona donde se encuentra la anotación con la zona que se obtuvo en la detección.

3. Resultados



Fig. 1.

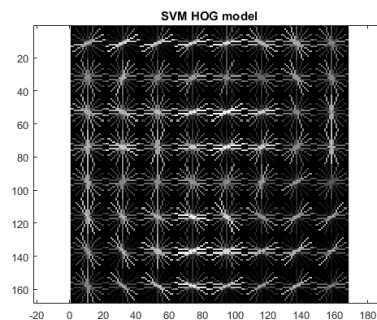


Fig. 2.

Como se puede observar, para el problema de detección de rostros lo que se tiene en principio es un modelo de HOG que toma las orientaciones que en promedio tiene una cara de frente. De esta manera, para los ejemplos positivos se tienen resultados como el que se puede apreciar en la figura 2 y con ellos se entrena el SVM. En el caso de los ejemplos negativos, la representación en HOG si difiere bastante, pues las orientaciones en toda la imagen son diferentes. Como resultado de detección de peatones se presenta un rendimiento muy bajo, en parte puede deberse a las imágenes utilizadas en la sección de Hard Negative Mining las cuales no presentaron una mejora significativa.

3.1. Limitaciones

Algunas de las limitaciones que presenta este tipo de clasificación son:

- La cantidad de imágenes de entrenamiento, puesto que aumentar el número de imágenes demasiado grande aumenta significativamente el costo computacional.

- Debido al uso de el algoritmo de Hard Negative Mining, el modelo obtenido puede variar significativamente dado a que los ejemplos utilizados para mejorar la clasificación son elegidos de manera aleatoria.
- El tamaño de los objetos de detección, tanto muy grandes como muy pequeños que no logran ser representados en la cantidad de escalas elegidas, lo que se traduce en que no se presenta una detección de las caras en las diferentes imágenes cuando estas son muy grandes o muy pequeñas.

3.2. Mejora

Se plantean las siguientes técnicas con el objetivo de mejorar la detección:

- El uso de otro tipo de descriptores de la imagen, como por ejemplo el uso del descriptor SURF.
- Con el uso de otros descriptores es posible crear un modelo basado en adaboost, es decir, que cada uno de los descriptores den como respuesta una hipótesis débil de clasificación para que al tener todas las respuestas de los diferentes clasificadores logre obtenerse un clasificador fuerte que tenga un mejor desempeño para las imágenes.

4. Bibliografía

1. Histogram of Orientation Gradients. School of Computer Science. Dublin. Ireland
2. Joao Henriques, Joao Carreira, Rui Caseiro, Jorge Batista. Beyond Hard Negative Mining. Universidad de Coimbra.
3. Chung Chang y Jen Lin (2001) A Library for Support Vector Machines. Universidad nacional de Taiwan.