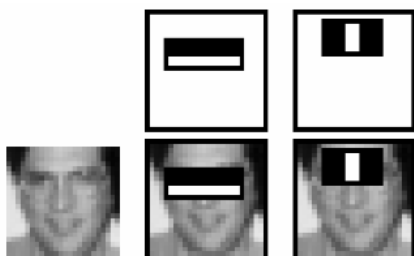


Implementación Integral Channel Feature

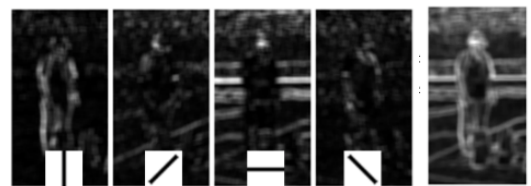
Pablo Fierro

En la tarea de detección de peatones existen diversas técnicas que se pueden utilizar, en este experimento usamos una variante del algoritmo propuesto por Viola y Jones en el 2001 [1], extiende la idea de los descriptores Haar y lo aplica a una serie de imágenes calculadas a partir de la imagen original, el algoritmo utilizado en este experimento se conoce como Integral Channel Features [2], para entender este algoritmo hay que saber sobre que extiende y como funciona el algoritmo de Viola y Jones y los descriptores utilizados en el:

Los descriptores Haar utilizan regiones rectangulares (y otras formas en algunas variaciones de la implementación) para calcular diferencias de intensidad en una imagen dada, luego estas regiones son usadas para identificar cualidades de objetos en una imagen, por ejemplo: en una cara humana hay patrones en las variaciones de intensidad, una de ellas es la region de la nariz y los ojos, los ojos son mas claros que la nariz, esta es una de las cualidades que Haar describe acerca de una cara humana, en el algoritmo de detección se utilizan una serie de regiones que describen el objeto deseado y se entrena un clasificador para identificar estas regiones en otras imágenes. Lo que el clasificador recibe a partir de estos descriptores es una suma de todos los pixeles en la geometría de cada descriptor.



Una vez descrito el algoritmo en forma general en el cual se basa el algoritmo de Integral Channel Features, podemos decir que este algoritmo funciona de la misma forma con algunas variaciones: en vez de utilizar la suma de intensidades de una misma imagen, generamos diferentes tipos de imágenes conocidas como canales, cada canal proviene de la imagen original y pueden ser: imagen a escala de grises, gradiente, histograma de orientaciones del gradiente, diferentes espacios de colores, etc.



El descriptor utilizado en este algoritmo de detección utiliza la suma sobre una imagen integral generada de cada canal, la imagen integral es la suma en cada pixel de todos los pixeles anteriores. Cada descriptor se compone de una geometría rectangular generada al azar dentro de las dimensiones de la ventana con la cual entrenaremos el clasificador, en el experimento utilizamos ventanas de 96x160 que provienen del dataset INRIA, además de la geometría, se debe seleccionar para cada descriptor el canal a utilizar.

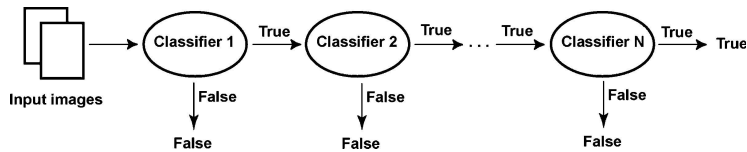
Cada descriptor genera la suma de la imagen integral de su canal respectivo sobre su geometría. Por ejemplo,



en la siguiente imagen el descriptor utiliza el canal gradiente y su geometría está marcada con verde, el resultado de este descriptor es la suma de pixeles sobre la imagen integral de este canal.

Para el experimento se utilizaron 10,000 descriptores.

El clasificador del algoritmo de detección se conoce como Adaboost, es una combinación lineal de clasificadores asignados un peso cada uno, para el experimento utilizamos arboles de decisión de máximo 2 niveles con un total de 1,000 arboles.

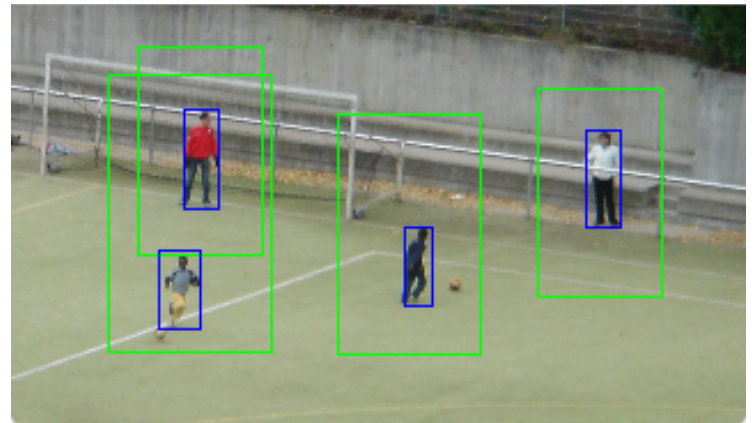


Los descriptores utilizados en el clasificador van tomando cierta importancia en cada nivel del clasificador, esto separa de que descriptores o features de la imagen son más importantes para la detección de peatones.

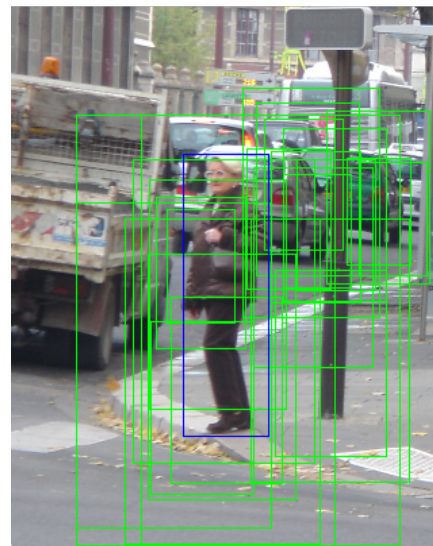
Una vez entrenado el clasificador, la detección en imágenes reales se aplica barriendo una ventana de la misma dimension con la cual se entreno — 96x160 en este experimento— sobre toda la imagen desplazándola con un valor d , en cada desplazamiento se genera el resultado de todos los descriptores utilizados para el entrenamiento y se evalúa el clasificador con esos datos, si el clasificador regresa un valor asociado con la clase persona, en la posición actual de la ventana se encuentra un posible candidato, este proceso se realiza sobre toda la imagen y a su vez escalando la imagen en diferentes tamaños, más grande y más pequeña.

Resultados

No fue posible obtener los mismos resultados obtenidos en [2] debido a que el entrenamiento y velocidad de obtención de los resultados era muy largo y no se pudo probar diferentes configuraciones para lograr resultados similares sobre el data set INRIA, algunos ejemplos:



Los recuadros verdes son los obtenidos utilizando el algoritmo de detección y los recuadros azules son los recuadros esperados, como podemos ver en esta imagen obtenemos el 100% de clasificación correcta, sin embargo el resultado no se repite para otras imágenes.



Referencias

- [1] Viola, Jones. (2005). Detecting Pedestrians Using Patterns of Motion and Appearance. Springer Science + Business Media, Inc.
- [2] Dollár. et. al (2009). Integral Channel Features.