

# HOG Face Detection

José María Campo Viñas  
Universidad de los Andes  
Código: 201412002

jm.campo11@uniandes.edu.co

Mariana Franky  
Universidad de los Andes  
Código: 201313944

m.franky10@uniandes.edu.co

## Abstract

*En este laboratorio se utilizó HOG como descriptor de características con el objetivo de hacer un detector de caras. Para esto se utilizó un subconjunto de la base de datos WIDER FACES, se utilizó como código base el código creado por el grupo de Geometría Visual de Oxford.*

## 1.. Introducción

Histograma de gradientes orientados es un descriptor de características que divide la imagen en bloques distribuidos a lo largo y ancho de la imagen, cada uno de los bloques se subdivide  $N \times M$  celdas y en cada una de ellas se calcula la magnitud y orientación de los gradientes. Para cada una de las celdas se calcula un histograma con todas las orientaciones detectadas y se forma el vector característico de la imagen. Este ha sido utilizado en los problemas de detección debido a que este detecta la forma del objeto mediante los gradientes orientados, haciendo que sea invariante a transformaciones geométricas y fotométricas.[1][2]

Entre los hiperparámetros que se pueden encontrar está el tamaño de la ventana, este es útil para definir el tamaño de nuestro objeto de interés. El tamaño de la celda es útil para definir el número de píxeles que se requiere por celda, cuando la celda es muy grande no se puede tener una descripción local de la imagen y cuando es demasiado pequeño se suprime información espacial valiosa. Número de bins: Un bin es la presentación de cuantificación de un espacio. Generalmente, en el bin se cuenta la cantidad de elementos del espacio que existen en la región definida por dicho bin.[1][2]

## 2.. Métodos y materiales

### 2.1.. Materiales

En este laboratorio se utilizó los recursos de Multimedia Laboratory de la universidad China de Hong Kong. Para abordar el problema se utilizó un subconjunto de la base de datos WIDER FACES, sólo se trabajó con instancias

de cara cuyas dimensiones sean mayores que 80 X 80 píxeles, así se evitó aquellas imágenes que eran bastante pequeñas o tienen una cantidad significativa de desenfoque. Para el entrenamiento de falsos negativos se empleó la clase BACKGROUND\_Google perteneciente a la base de datos de Caltech-101, de ésta se eliminaron todas las imágenes que mostraran rostros, un ejemplo se muestra en la Figura 1. Mientras que para el entrenamiento de verdaderos positivos se empleó la subclase TrainCrops de la base de datos WIDER\_FACE.

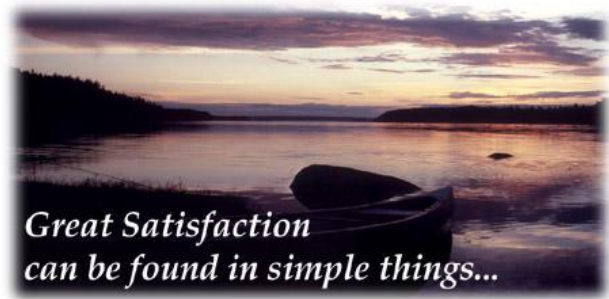


Figura 1. Muestra de falsos negativos.

### 2.2.. Metodología:

El código base que se utilizó para este laboratorio fue creado por el grupo de Geometría Visual de Oxford. Se tomó como base el ejercicio 5, en el que uno puede entrenar su propio detector. A continuación se desarrollará una breve descripción de nuestro método:

- Inicialmente se inicializan los parámetros:  $\text{hogCellSize} = 8$ ;  $\text{minScale} = -1$ ;  $\text{maxScale} = 3$ ;  $\text{numOctaveSubdivisions} = 3$ ; Estos parámetros fueron tomados del algoritmo base. Las modificaciones que se realizaron en este código fue al construir la base de datos negativos y positivos.
- Extracción de características: para extraer las características de la imagen se utiliza la función de `vl_hog`.

- Se inicializa los datos positivos y negativos y se entrena el modelo: en la cual se entrena los modelos con un SVM mediante la función `v_svmtrain`, además se creará un vector de etiquetas binarias, donde 1 equivale a los puntos positivos y -1 para los puntos negativos. Se utilizó el parámetro C para SVM.
- Una vez entrenado el modelo se evalúa el modelo los datos de entrenamiento y además se detectan los hard negatives. Una vez detectados los hard negatives, se eliminan aquellos que están duplicados.
- Ahora bien, para evaluar nuestro modelo no se utilizó el código estándar de evaluación debido a que es un subconjunto muy pequeño de los datos original. Se utilizó una versión modificada de la versión del código. Este evalúa la intersección de la unión del bounding box del ground truth con el bounding box del resultado.

### 3.. Resultados

En la Figura 2 se muestra el histograma de gradientes orientados resultante del entrenamiento de rostros junto al Hard Negative Mining.

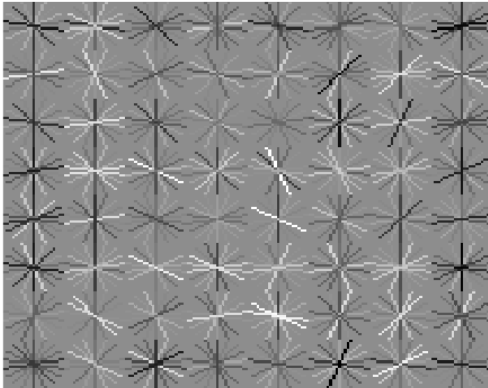


Figura 2. Característica HOG resultante del entrenamiento.

Los resultados de F-Medida se encuentran en el Cuadro 1 a continuación.

Dificultad	F-medida
Fácil	0.051
Medio	0.027
Difícil	0.011

Las curvas de precisión-cobertura para la evaluación con rostros a reconocer según los niveles de dificultad se encuentran a continuación en las Figuras 3 4 5.

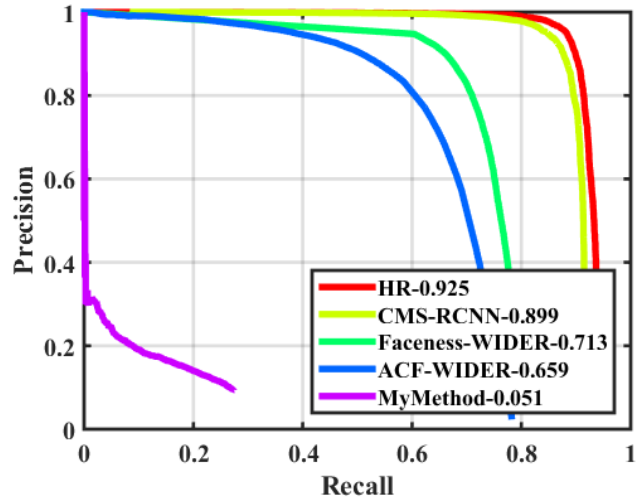


Figura 3. Curva precisión y cobertura para reconocimiento de rostros de dificultad fácil.

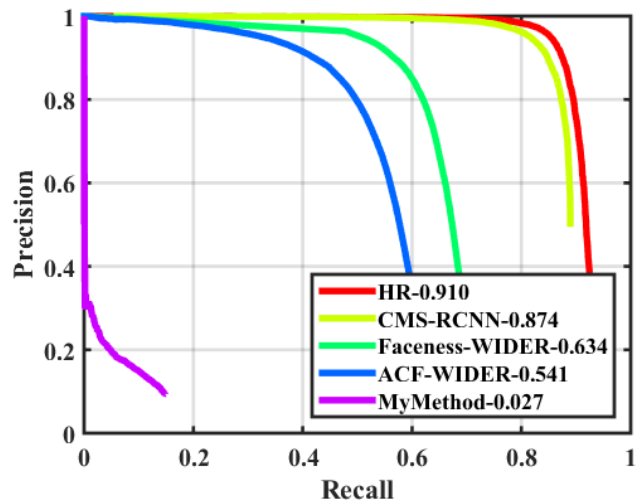


Figura 4. Curva precisión y cobertura para reconocimiento de rostros de dificultad medio.

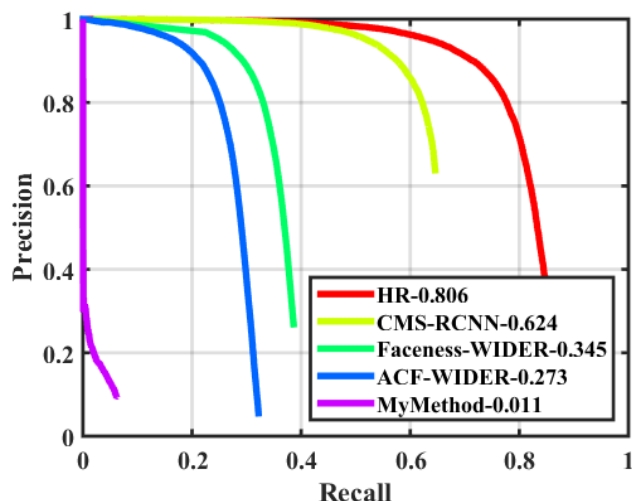


Figura 5. Curva precisión y cobertura para reconocimiento de rostros de alta dificultad.

#### 4.. Discusión y Conclusiones

Una de las mayores limitaciones en uso de HOG para clasificación de imágenes, es la presencia de objetos con una estructura similar al objeto que se desea buscar, de manera que si se busca un balón de fútbol, éste puede ser confundido con uno de básquetbol, ya que su diferencia radica en los colores.

Otra limitación es la definición de número de celdas y escalamiento de la imagen, si bien escogemos un número de celdas alto, tomará hasta los más mínimos detalles de una imagen, e ignorará la estructura que lo conforma.

Los falsos positivos presentan un patrón repetitivo, el cual consiste en que su forma externa es cercana a una elipse, esto se debe a que es la misma forma externa presenta en las imágenes de rostros. Las estructuras imagen interna de los objetos en la imagen no son tenidas en cuenta para la detección de rostros debido a que las imágenes de rostros en entrenamiento se encuentran en distintas poses, lo cual imposibilita su alineamiento.

Un ejemplo de patrones en Falsos Positivos se muestra en la Figura 6, donde ésta presenta gradientes fuertes en la forma externa del objeto y poseen una conformación de elipse.

Las imágenes de Falsos Negativos no presentan un patrón reconocible a primera instancia, ya que fueron extraídas aleatoriamente del buscador Google, lo anterior es lo que se buscaba, de manera que el clasificador aprendiese a lidiar con el abarrotamiento del fondo en las imágenes.

Una posible mejora al método, sería la creación de rostros categorizadas en la dirección a la cual se encuentra

mirando, de manera que el problema ya no sería binario, sino de multiclase, pero se lograría una mejor alineación entre imágenes, lo cual conllevaría a un histograma de bordes más definidos.



Figura 6. Muestra de Falsos Positivos.

#### 5.. Extra Credit

Waldo fue encontrado en la clase 13–Interview, en la imagen 13\_Interview\_Interview\_On\_Location\_13\_559.jpg. Éste se muestra en la Figura 7.

Waldo fue encontrado usando su HOG como modelo de entrenamiento y reescalándolo a los valores de 260x177, luego se escogieron por rectificación humana las imágenes con mayor puntaje de clasificación.



Figura 7. Waldo.

## Referencias

- [1] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 886–893, IEEE, 2005.
- [2] S. Ruiz, M. Rigoberto, T. Yagos, and W. Rolando, “Análisis de métodos de reconocimiento de rostros utilizando descriptores sift y hog aplicado en la escuela nicanor larrea,” B.S. thesis, Escuela Superior Politécnica de Chimborazo, 2015.