Master in
Computer Vision
Barcelona

INTRODUCTION TO HUMAN AND COMPUTER VISION 2017

TRAFFIC SIGN DETECTION / RECOGNITION

Àlex Vicente Sola
Àlex Palomo Domínguez
Joan Sintes Marcos
Roger Marí Molas

Task 1. Data analysis






Signal type	A	B	C	D	E	F
Max size	18502 pixels	9165 pixels	43809 pixels	31341 pixels	37035 pixels	52689 pixels
Min size	459 pixels	1157 pixels	1084 pixels	720 pixels	1066 pixels	1850 pixels
Form factor	1.0625	1.0178	0.9487	0.9744	0.9423	0.8268
Filling ratio	0.5007	0.4961	0.7839	0.7793	0.7846	0.9993
Frequency of appearance	0.2621	0.0356	0.1196	0.1807	0.0967	0.3053

Very indicative of the signal shape!!!

Group signals

SHAPE

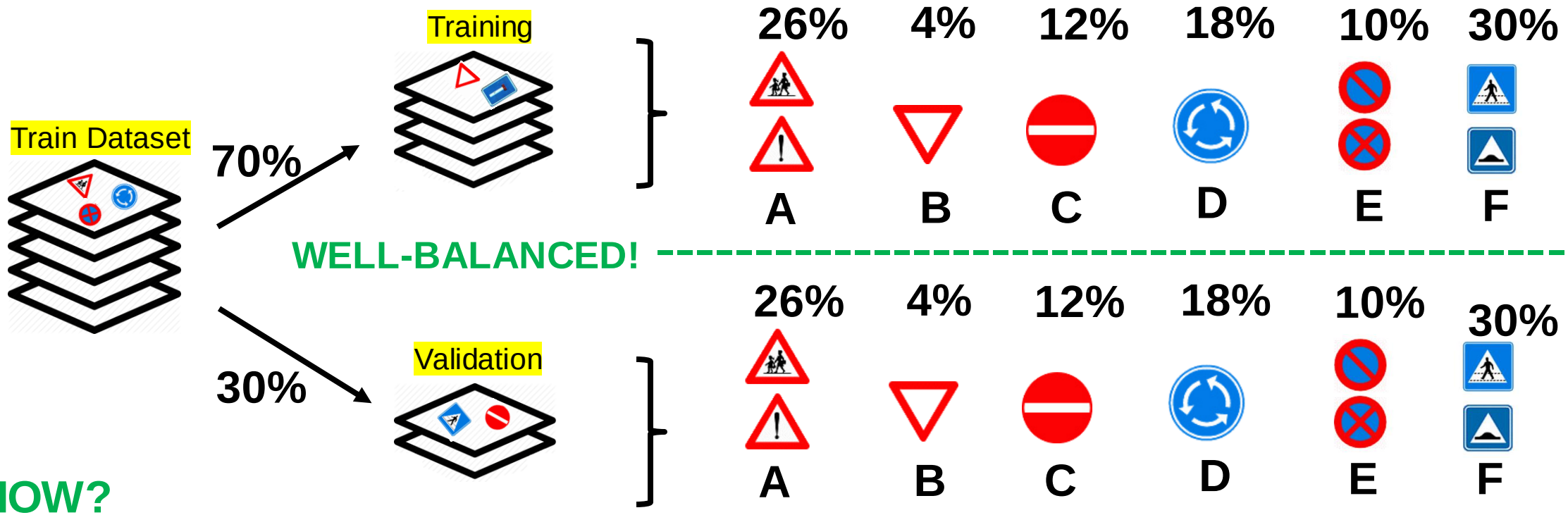
- Triangles: **A, B**
  
- Circles: **C, D, E**
  
- Quadrilaterals: **F**
  

COLOR (*)

- Red: **A, B, C**
  
- Blue: **D, F**
  
- Red + Blue: **E**
  

(*) Groups at first sight. Histograms (task 3) can be used to get a better idea about the real degree of similarity.

Task 2. Create Train and Validation Sets



HOW?

while training set is not complete

pick X images / signal type ← **Randomly!**

*where X = freq. of appear. * amount of images left to pick*

remove repeated images (*)

end

(*) Some images contain more than one type of signal
→ **Preserve only non-repeated images!**

(The images that are not part of the training set form the validation set)

Task 3. Mask Generation via Color Segmentation

YCbCr Color Space

Red

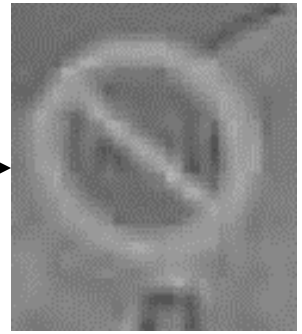
(RED MASK)

2 MASKS

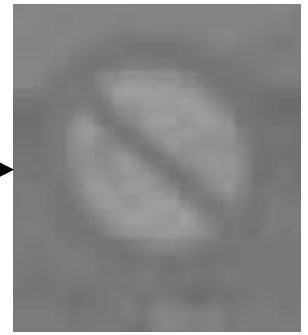
Blue

(BLUE MASK)

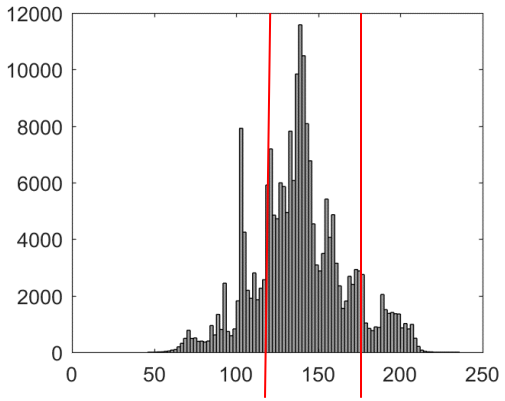
***Cr* channel**



***Cb* channel**

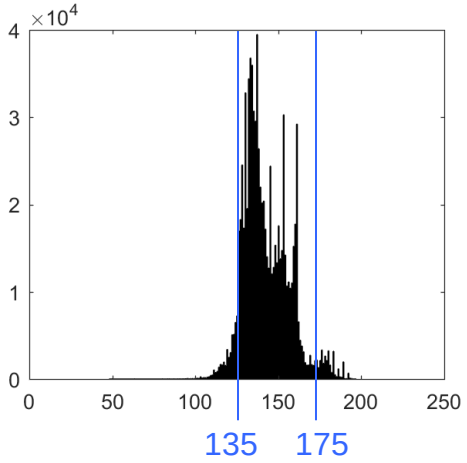


***Cr* channel *histogram* (*)**



*(*) from all pixels of the **A**, **B**, **C** signals*

***Cb* channel *histogram* (**)**

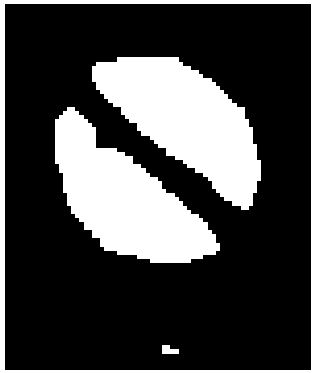


*(**) from all pixels of the **D**, **F** signals*

135* < *Cr* channel < *175

(RED MASK Cr)

135* < *Cb* channel < *175



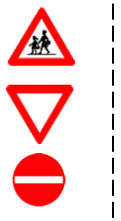
(BLUE MASK Cb)

Hand picked thresholds based on histograms!

Task 3. Mask Generation via Color Segmentation

HSV Color Space

Red



(RED MASK)

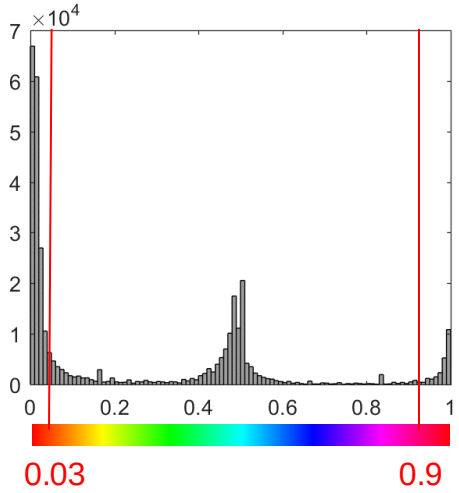
2 MASKS

Blue

(BLUE MASK)

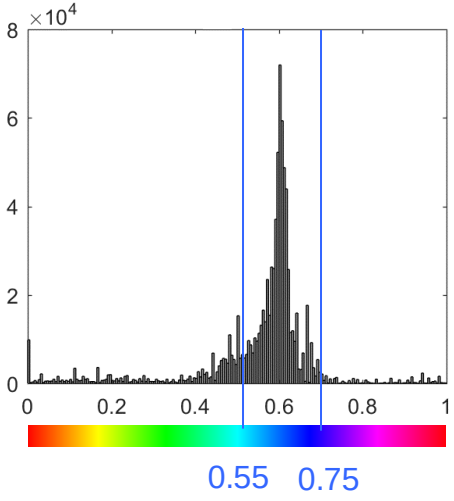
***Hue** channel*

***Hue** channel **histogram** (*)*



*(*) from all pixels of the **A**, **B**, **C** signals*

***Hue** channel **histogram** (**)*

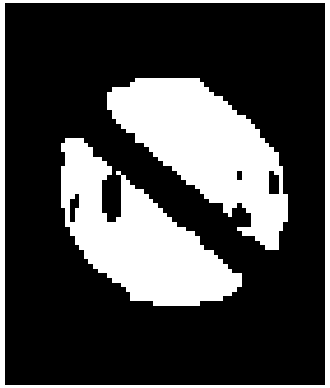


*(**) from all pixels of the **D, F** signals*

Hue** channel < **0.03** **OR** **Hue** channel > **0.9

(RED MASK H)

0.55** < **Hue** channel < **0.75



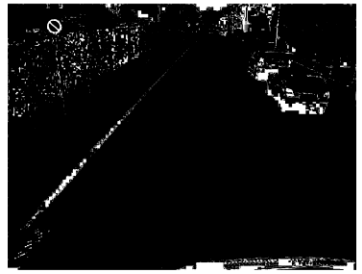
(BLUE MASK H)

Hand picked thresholds based on histograms!

Task 3. Mask Generation via Color Segmentation

RED MASK Cr

AND



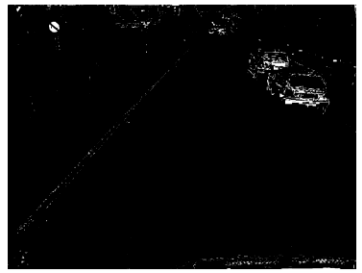
RED MASK H

=

RED MASK

BLUE MASK Cb

AND



BLUE MASK H

=

BLUE MASK

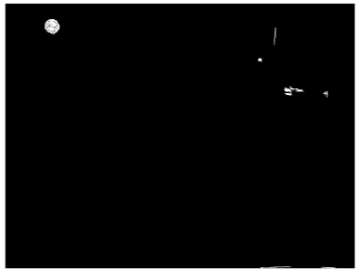
+

COMBINED MASK

Problem 1: Existence of white noise or large areas (due to red cars, etc.) that are not signals

Problem 2: Existence of black holes inside the detected signals

To minimize Problem 1:



COMBINED MASK
+
Size Filtering

*(Connected components of the mask smaller than **200px** or bigger than **60.000px** are filtered (*))*

Computationally expensive!

To minimize Problem 2:

COMBINED MASK
+
Size Filtering
+
Hole Filling

(Holes of the connected components of the mask are filled)

(*) Thresholds based on the **min** and **max** signal size from the Data Analysis (Task 1).

Task 4. Evaluation with Ground-Truth

We can expect the Accuracy to be high due to the large amount of TN, since the majority of pixels are not part of signals (value 0 in the mask). Not a very indicative measure in this context.

Harmonic mean of Recall and Precision! Very indicative!

Size Filtering is computationally expensive and increases dramatically the time per frame

Methods	Precision	Accuracy	Recall	F1-mesure	TP	FP	FN	Time per frame
Hue	0,034	0,903	0,585	0,061	$4{,}79 \cdot 10^3$	$1.91 \cdot 10^5$	$3.00 \cdot 10^3$	0,14 s
CbCr	0,113	0,963	0,532	0,157	$4{,}39 \cdot 10^3$	$7{,}18 \cdot 10^4$	$3{,}40 \cdot 10^3$	0,13 s
CbCr + H	0,266	0,989	0,448	0,284	$3{,}83 \cdot 10^3$	$1{,}84 \cdot 10^4$	$3{,}97 \cdot 10^3$	0,16 s
CbCr + H + Size Filtering	0.449	0.994	0.421	0.384	$4.18 \cdot 10^3$	$8.23 \cdot 10^3$	$4.61 \cdot 10^3$	2.58 s
CbCr + H + Size Filtering + Hole Filling	0,424	0,994	0,609	0,450	$5{,}41 \cdot 10^3$	$1{,}04 \cdot 10^4$	$2{,}39 \cdot 10^3$	2,93 s

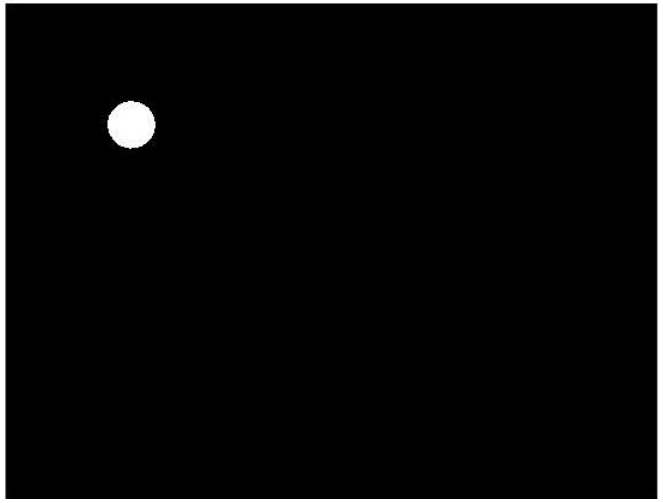
- **RECALL** measures the ratio of detected signals over the total amount of signals in the validation set.

  → It is a **crucial measure**: if a signal is missed in the detection step, then there is no chance to recognize it in the later steps.
  → We can tolerate having a certain amount of FP (which can be discarded later) as long as we reach the maximum possible TP.

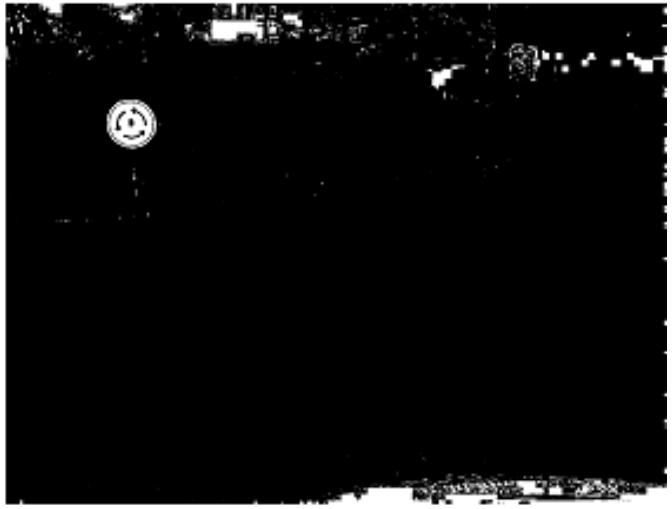
- **Best result (highest RECALL and F1-MEASURE) obtained in the 5th scenario.**

Task 4. Evaluation with Ground-Truth

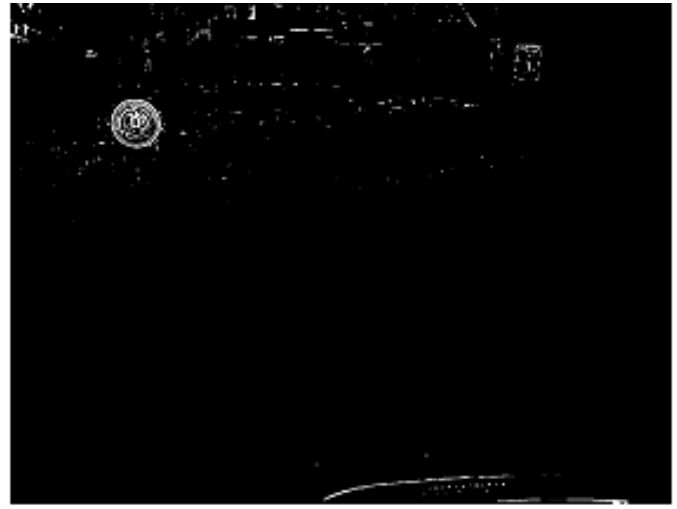
Ground-Truth mask



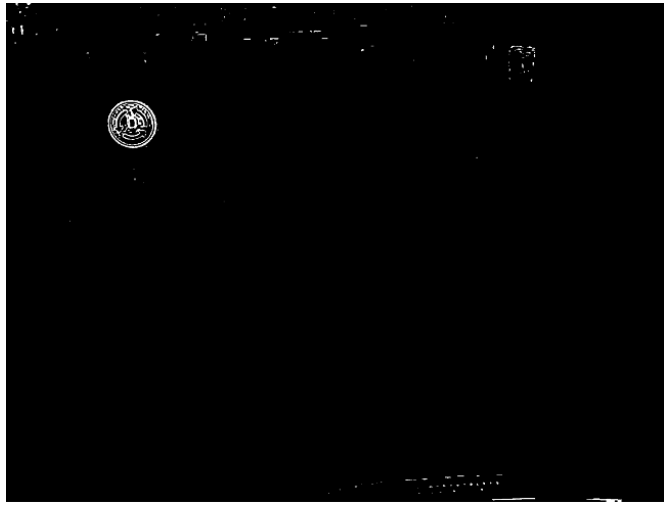
Hue



CbCr



CbCr + Hue



CbCr + Hue + Size filtering

CbCr + Hue + Size filtering + Hole filling