



# Extractor de subtítulos incrustados en videos

Trabajo de aplicación

Procesamiento Digital de Imágenes

2018

Arato, Nicolás  
Páez, Mariángeles  
Ramírez, Darién



# Subtítulos Incrustados

- Pegados a la imagen.
- No están en formato texto.



# Subtítulos Incrustados

- No se pueden editar.
- No se pueden reutilizar.
- No se puede acceder a información de los tiempos.
- No se pueden desactivar.



Propuesta:



# **Método automático de extracción de subtítulos incrustados en videos.**



# Etapas de detección de subtítulos

- 1 **Identificación de regiones**  
Se trabaja con la región inferior de la imagen correspondiente al 25% de la altura total de cada fotograma.
- 2 **Utilización del modelo HSV**  
Se utiliza el canal V.
- 3 **Borroneado de fondo**  
Filtro Pasa Bajo
- 4 **Detección de contornos de caracteres**  
Filtro Pasa Alto
- 5 **Binarización de la imagen**  
Umbral binario
- 6 **Eliminación de píxeles espúreos**  
Operaciones morfológicas - Filtro de mediana
- 7 **Detección de caracteres en la región**  
Transformada de Hough



# Etapas de detección de subtítulos

Tenía información crucial  
para la humanidad y no la compartió.

Tenía información crucial  
para la humanidad y no la compartió.

Tenía información crucial  
para la humanidad y no la compartió.

Tenía información crucial  
para la humanidad y no la compartió.

Tenía información crucial  
para la humanidad y no la compartió.

Tenía información crucial  
para la humanidad y no la compartió.

Tenía información crucial  
para la humanidad y no la compartió.

Tenía información crucial  
para la humanidad y no la compartió.



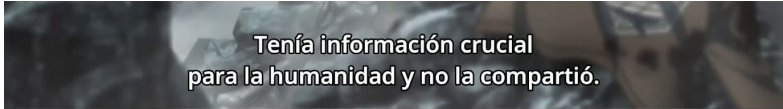
# Etapas de detección de subtítulos

Filtrado PB y  
PA

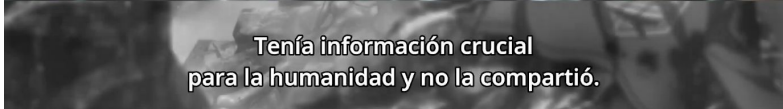
Umbral binario

Op morfológicas  
y mediana

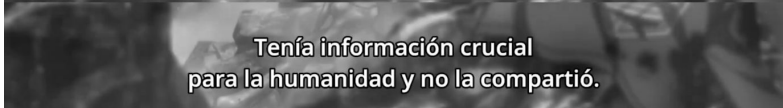
Hough



Tenía información crucial  
para la humanidad y no la compartió.



Tenía información crucial  
para la humanidad y no la compartió.



Tenía información crucial  
para la humanidad y no la compartió.



Tenía información crucial  
para la humanidad y no la compartió.



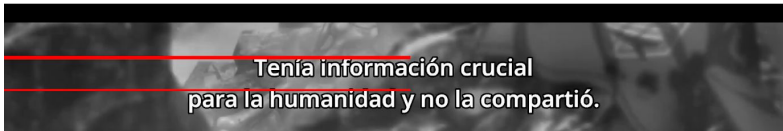
Tenía información crucial  
para la humanidad y no la compartió.



Tenía información crucial  
para la humanidad y no la compartió.



Tenía información crucial  
para la humanidad y no la compartió.



Tenía información crucial  
para la humanidad y no la compartió.



# Preprocesamiento de imagen para utilizar el detector OCR

- Promediado de fotogramas consecutivos con el mismo subtítulo.
- Transformación de potencia.

Es cierto que sus decisiones  
hasta ahora son imperdonables.



# TESSERACT OCR

- Licencia Libre
- Lista blanca

abcdefghijklmnopqrstuvwxyz  
vwxyz ABCDEFGHIJKLMN  
ÑOPQRSTUVWXYZ áéíóú  
ÁÉÍÓÚüÜ.,;!¿?





# Confección de archivo SubRip

Formato de archivo .srt

## **Número de subtítulo**

(En orden secuencial empezando en 1 para el primer subtítulo)

## **Tiempo inicial --> Tiempo final**

(En formato horas:minutos:segundos,milisegundos)

## **Texto del subtítulo**

(Puede incluir una o varias líneas separadas por un salto de línea)

## **Línea en blanco**



# Confección de archivo SubRip

Ejemplo

*1*

*00:00:00,394 --> 00:00:03,031*

*Anteriormente en*

*“Sons of Anarchy”*

*2*

*00:00:03,510 --> 00:00:05,154*

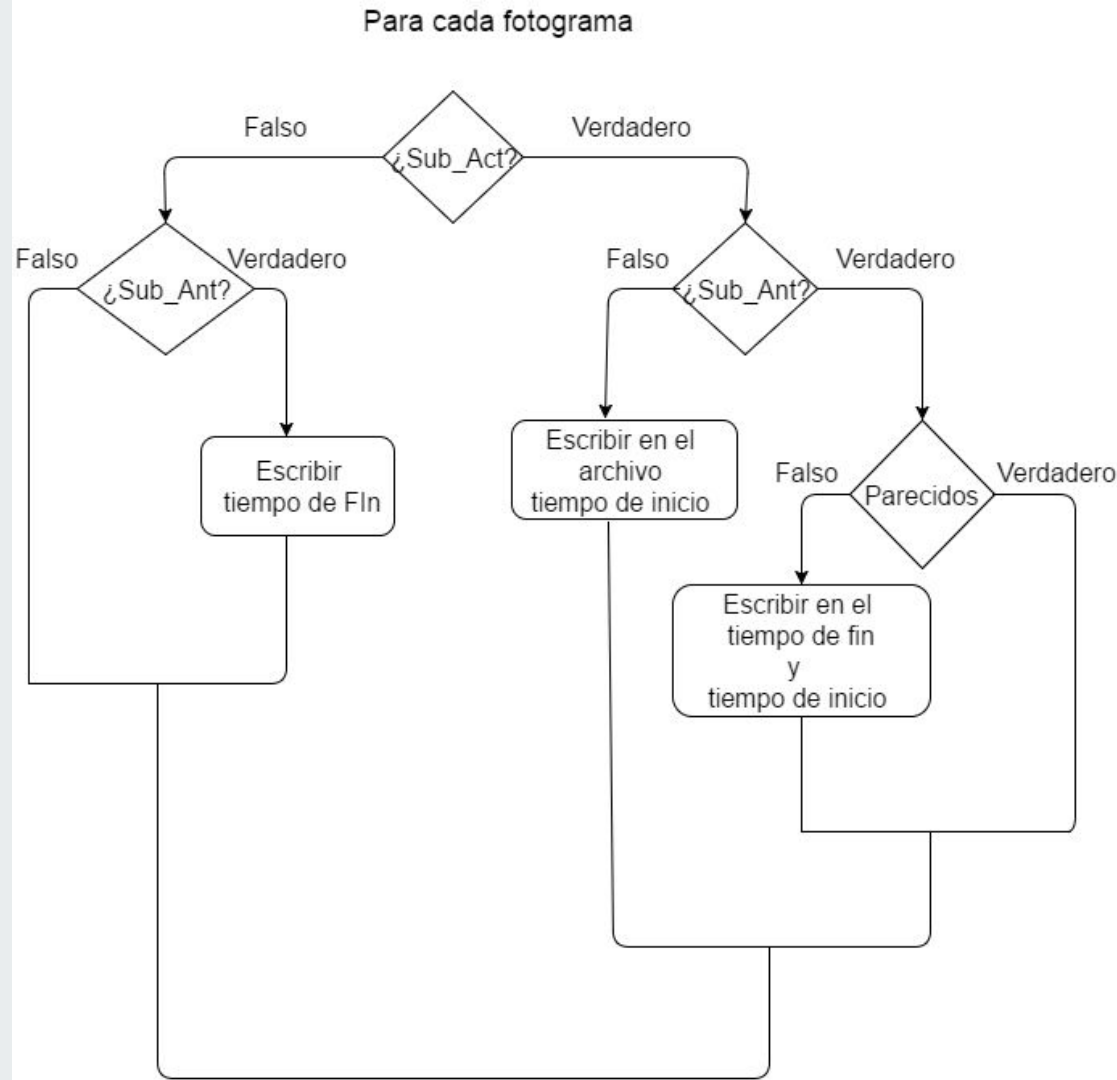
*Ponte esto. Ayudará.*

*3*

*00:00:05,274 --> 00:00:07,021*

*Estoy bien sin él.*

# Confección de archivo SubRip



# Medidas para la calidad de la detección de subtítulos

Predicción



Solución



$$Dice = \frac{2 |Solución \cap Predicción|}{|Solución| + |Predicción|} = \frac{2 TP}{FN + 2 TP + FP}$$

$$Jaccard = \frac{|Solución \cap Predicción|}{|Solución \cup Predicción|} = \frac{TP}{FN + TP + FP}$$



# Resultados

Archivo	Resolución	Dice	Jaccard
snk[1]	720 (HD)	0,980735	0,962477
snk[11]	720 (HD)	0,994318	0,988724
snk[20]	720 (HD)	0,889938	0,805134
snk[1]	1080 (FullHD)	0,979918	0,960946
snk[11]	1080 (FullHD)	0,994318	0,988724
snk[20]	1080 (FullHD)	0,889938	0,805134
kf[6]	720 (HD)	0,981076	0,963177
kf[10]	720 (HD)	0,948524	0,903263
kf[12]	720 (HD)	0,953312	0,912434
kf[6]	1080 (FullHD)	0,982348	0,96558
kf[10]	1080 (FullHD)	0,961848	0,92728
kf[12]	1080 (FullHD)	0,963206	0,929911
y[1]	480 (SD NTSC)	0,974711	0,951023
y[6]	480 (SD NTSC)	0,990482	0,981217
y[20]	480 (SD NTSC)	0,984751	0,970172



# Conclusiones

- Tiempo alto: dependiente de la resolución.
- No se detectan falsos negativos.
- Se detectan falsos positivos.
- Se detectan partes del fondo como subtítulos.
- Es sensible al ruido debido a la utilización de filtros de derivada segunda para la detección de bordes.
- El OCR detecta bordes residuales como caracteres válidos.



# Gracias