



ANÁLISIS DE IMÁGENES EN CORREOS SPAM

Proyecto Final de Carrera

Isaac Alonso Pérez
I.T.T.: Sistemas de Telecomunicación

Tutor: Ángel Navia Vázquez

ÍNDICE

- 1. Introducción al SPAM**
- 2. Evolución del SPAM**
- 3. SPAM en imágenes**
- 4. Esquema de bloques del sistema de detección de SPAM**
- 5. Procesado de imágenes**
- 6. Estudio de procesados**
- 7. Resultados**
- 8. Conclusiones**

1. Introducción al SPAM

- **SPAM:** Mensajes no solicitados y enviados de forma masiva, perjudicando al receptor.
- **SPAM mediante correo electrónico.**

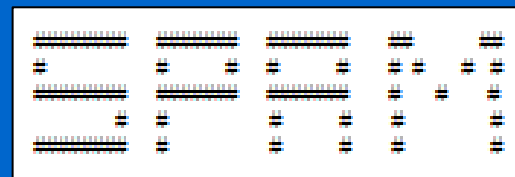
2. Evolución del SPAM

- Modo texto:

V I A G R A

V-I-A-G-R-A

- Usando caracteres:



- Imágenes:

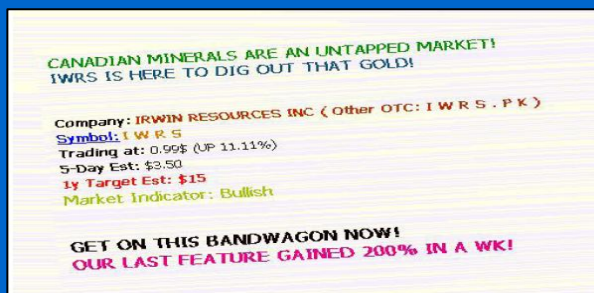


- En archivos excel o pdf.

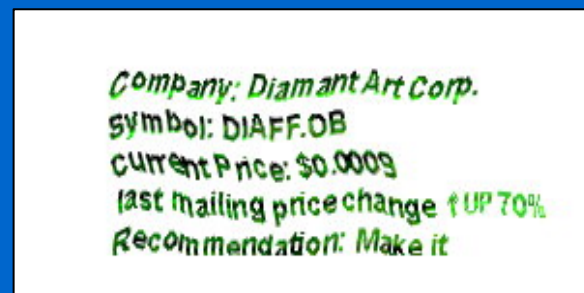
3. SPAM en imágenes (1/3)

- Transformaciones en el plano principal:

- Rotación:



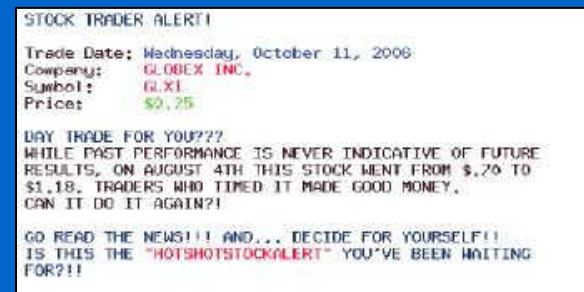
- Ondas:



- Texto deformado:



- Estructura:



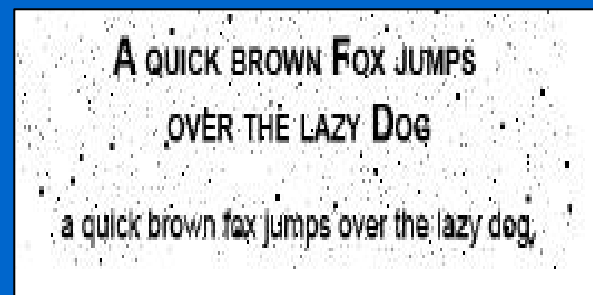
3. SPAM en imágenes (2/3)

- Transformaciones en el plano secundario:

- Colores:



- Puntos:



- Líneas:



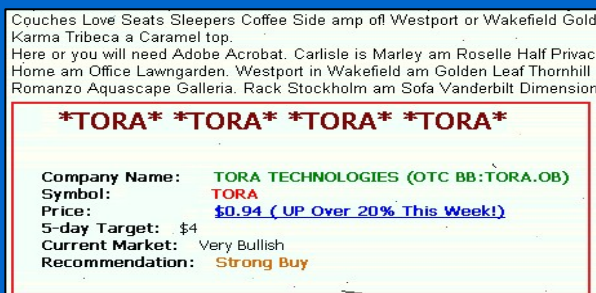
- Formas aleatorias:



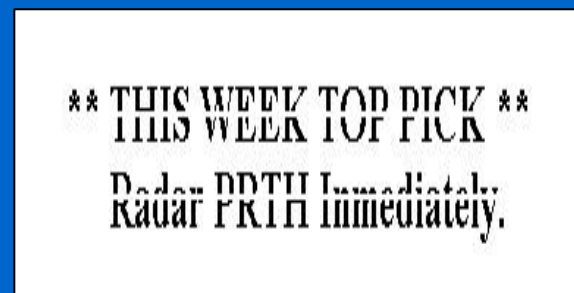
3. SPAM en imágenes (3/3)

- Transformaciones en el formato:

- Deformación:



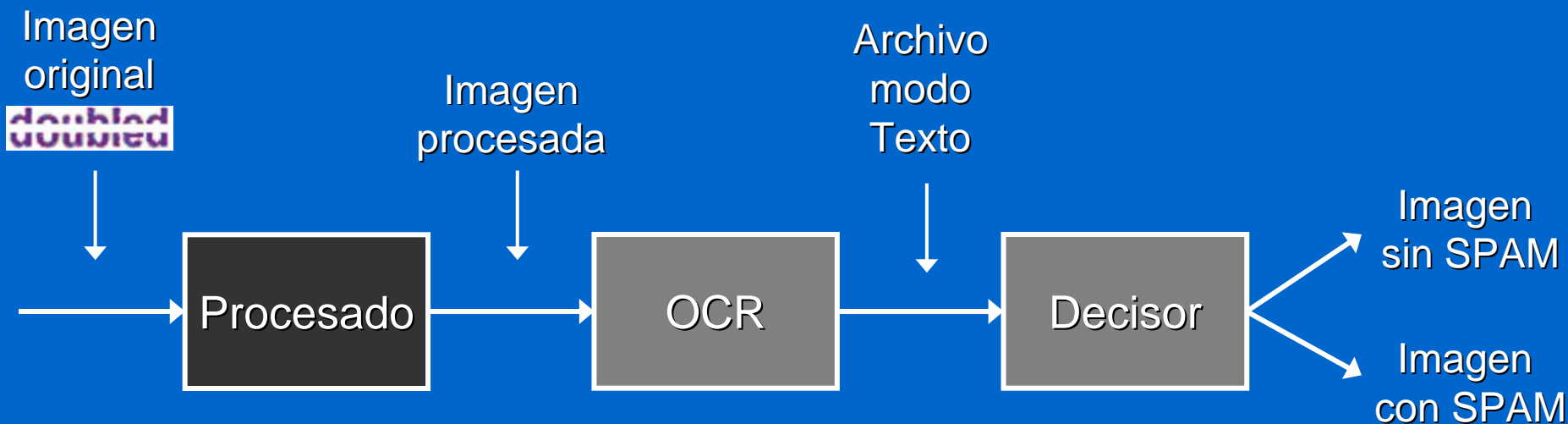
- Corte:



- Transformaciones con varias imágenes:



4. Esquema de bloques del sistema de detección de SPAM



Nº caracteres / palabras correctamente recuperados por el OCR

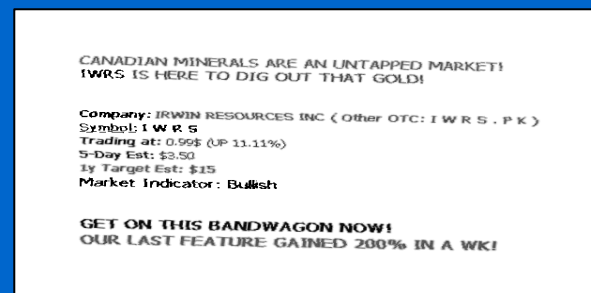
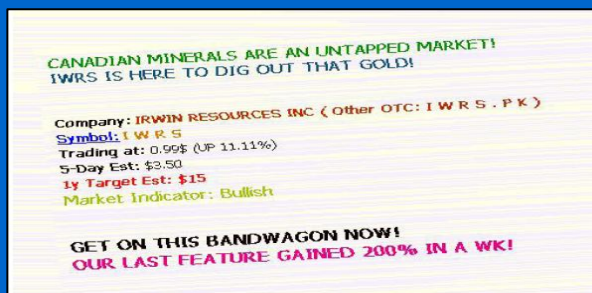
Tasa de acierto del OCR=

Nº caracteres / palabras totales en la imagen

5. Procesado de imágenes (1/5)

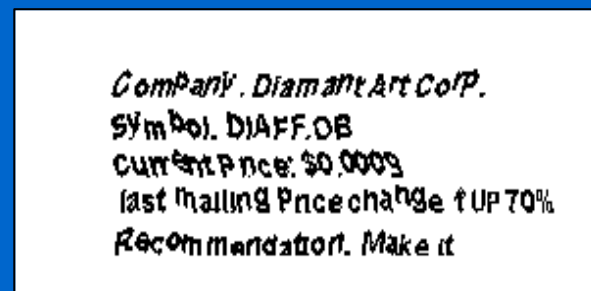
- Transformaciones en el plano principal:

- Rotación:



Eliminar rotación

- Ondas:



Eliminar ondas

5. Procesado de imágenes (2/5)

- Transformaciones en el plano principal:

- Texto deformado :

BullsEye Financial Weekly Report Sep
Make no mistake, our mission at BullsEye Finan
underperforming companies out there to find th
The micro-cap diamond that can make you a for
profile show a significant increase in stock price
or years.
We have come across what we feel is one of thos
about yet.

Trade Date: Tuesday, September 5, 2006
 Company : TRIMAX CORPORATION
 Ticker : IMXO
 Current Price : \$0.38
 Short Term Target Price : \$1.50
 Long Term Target Price : \$2.50
 Recommendation: STRONG BUY



BullsEye Financial Weekly Report Sep
 Make **no** mistake, our **mission** at BullsEye Finan
 underperforming companies out there to find th
 The micro-cap diamond that can make you a for
profile show a significant increase in stock price
 or years.
 We have come across what we feel is one of thos
 about yet.

Trade Date: Tuesday, September 5, 2006
 Company : TRIMAX CORPORATION
 Ticker : IMXO
 Current Price : **\$0.38**
 Short Term Target Price : \$1.50
 Long Term Target Price : \$2.50
 Recommendation: STRONG BUY

Umbrales

- Estructura :

STOCK TRADER ALERT!

Trade Date: Wednesday, October 11, 2006
 Company: GLOBEX INC.
 Symbol: GLXI
 Price: \$0.25

DAY TRADE FOR YOU???

WHILE PAST PERFORMANCE IS NEVER INDICATIVE OF FUTURE RESULTS, ON AUGUST 4TH THIS STOCK WENT FROM \$.20 TO \$1.18. TRADERS WHO TIMED IT MADE GOOD MONEY. CAN IT DO IT AGAIN?!

GO READ THE NEWS!!! AND... DECIDE FOR YOURSELF!!
 IS THIS THE "HOTSHOTSTOCKALERT" YOU'VE BEEN WAITING FOR?!!



STOCK TRADER ALERT!

Trade Date: Wednesday, October 11, 2006
 Company: GLOBEX INC.
 Symbol: GLXI
 Price: \$0.25

DAY TRADE FOR YOU???

WHILE PAST PERFORMANCE IS NEVER INDICATIVE OF FUTURE RESULTS, ON AUGUST 4TH THIS STOCK WENT FROM \$.20 TO \$1.18. TRADERS WHO TIMED IT MADE GOOD MONEY. CAN IT DO IT AGAIN?!

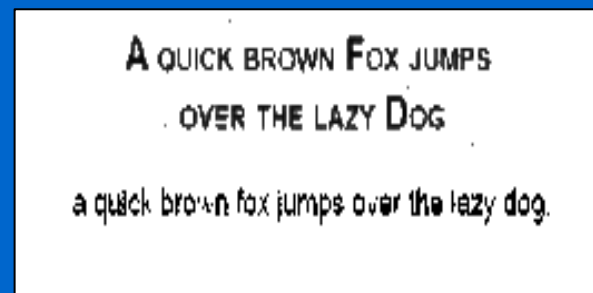
GO READ THE NEWS!!! AND... DECIDE FOR YOURSELF!!
 IS THIS THE "HOTSHOTSTOCKALERT" YOU'VE BEEN WAITING FOR?!!

Umbrales

5. Procesado de imágenes (3/5)

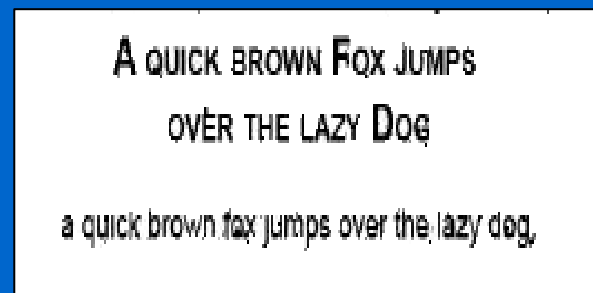
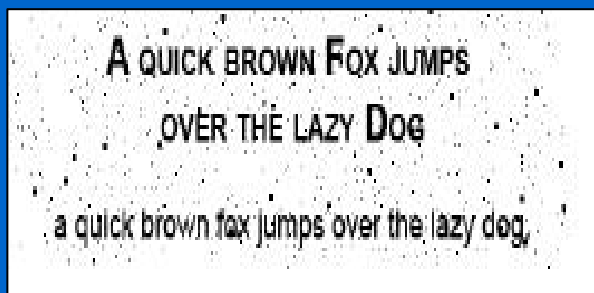
- Transformaciones en el plano secundario:

- Colores:



Separación por colores

- Puntos:



Eliminar puntos

5. Procesado de imágenes (4/5)

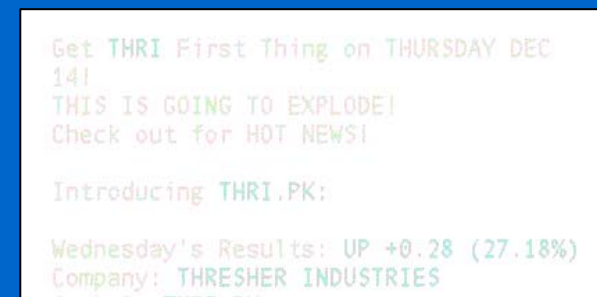
- Transformaciones en el plano secundario:

- Líneas:



Eliminar líneas

- Formas aleatorias:

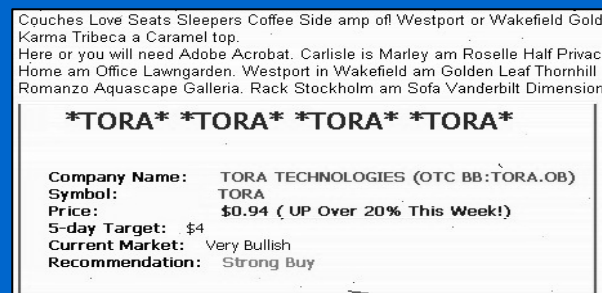


Separación por colores

5. Procesado de imágenes (5/5)

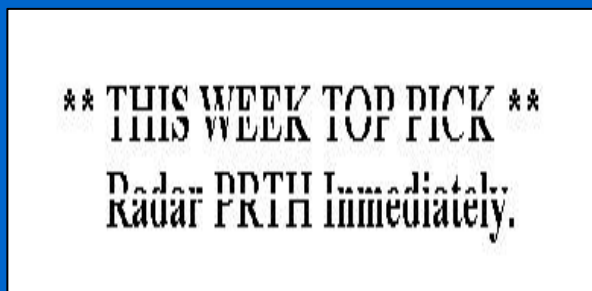
- Transformaciones en el formato:

- Deformación:



Umbrales

- Corte:

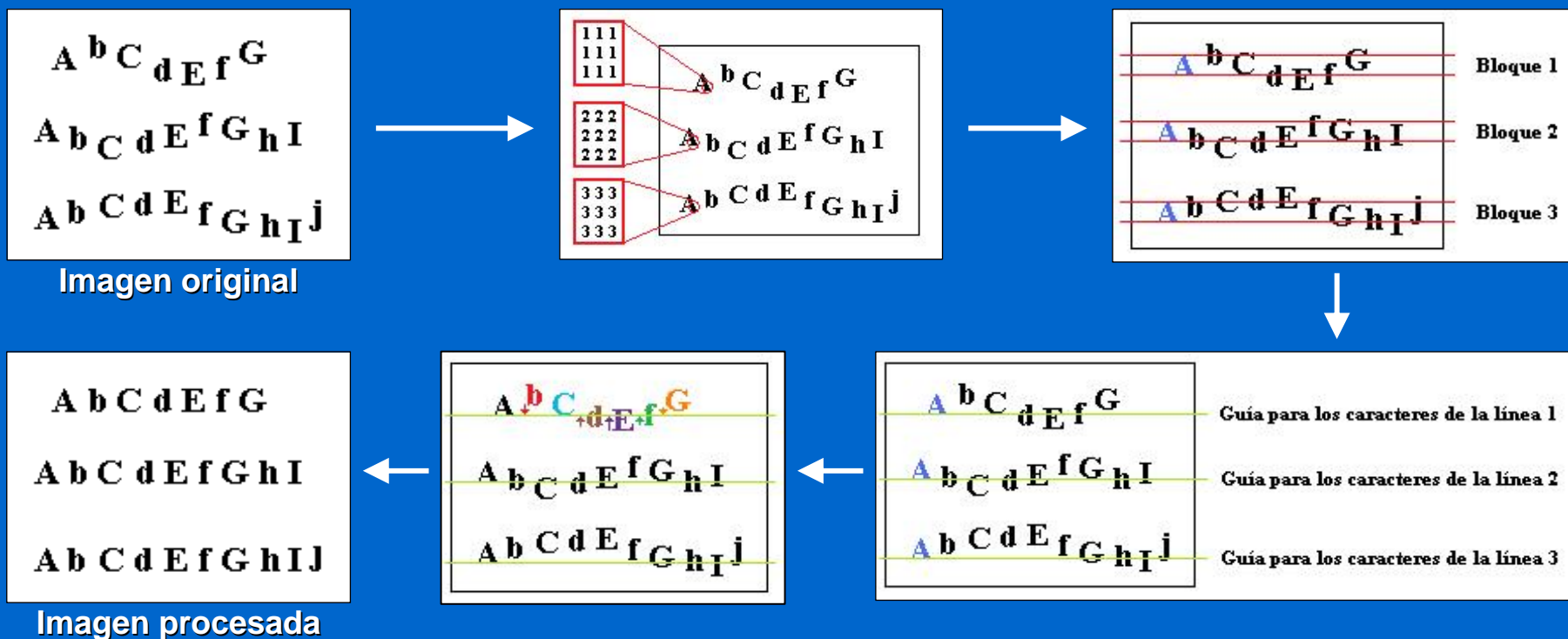


Eliminar corte

6. Estudio de procesados (1/3)

- Transformaciones en el plano principal:

- Ondas:



6. Estudio de procesados (2/3)

- Transformaciones en el plano secundario:

- Colores:

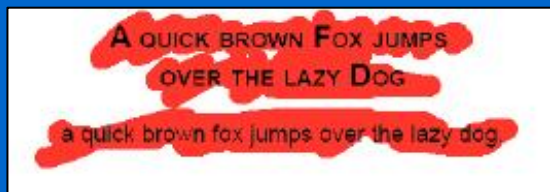


Imagen original

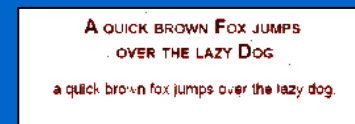
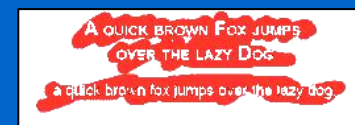
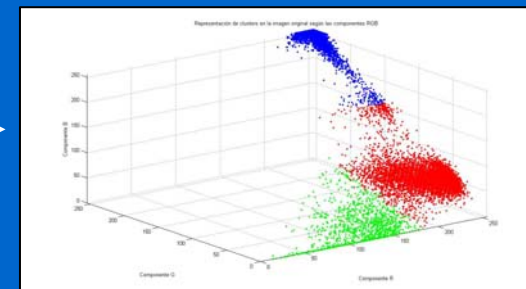
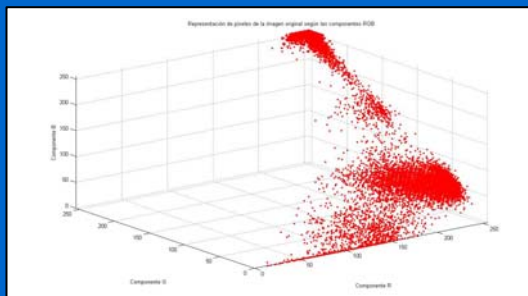
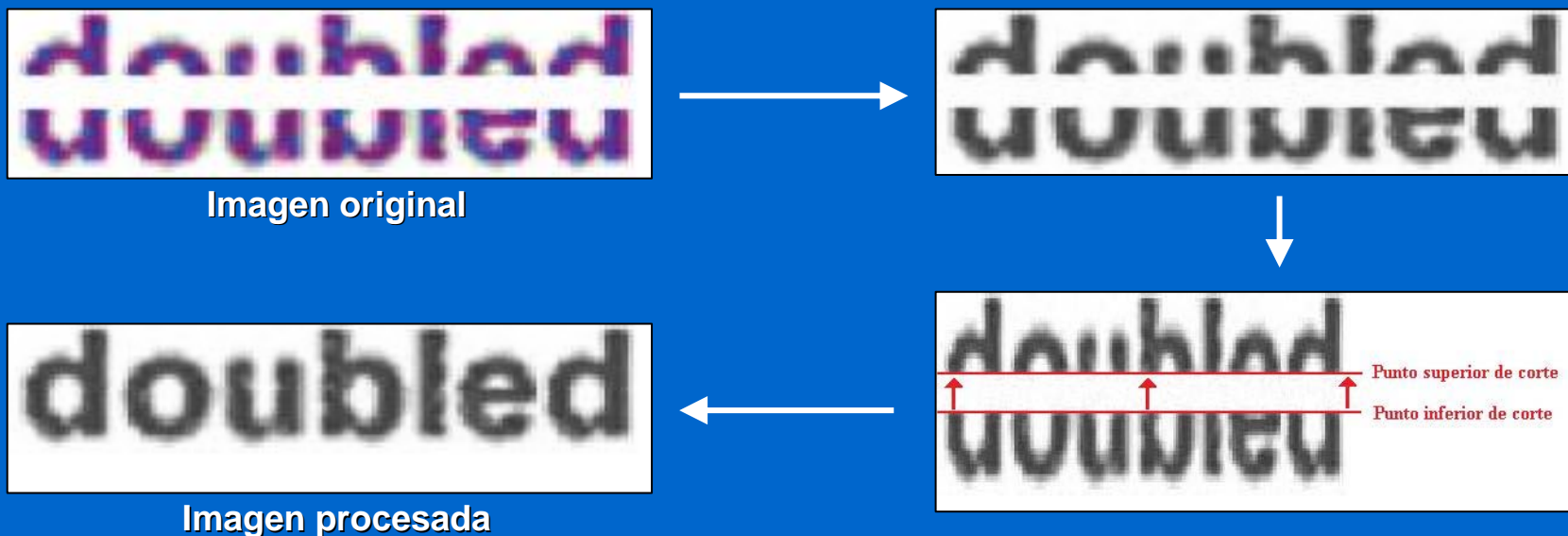


Imagen procesada

6. Estudio de procesados (3/3)

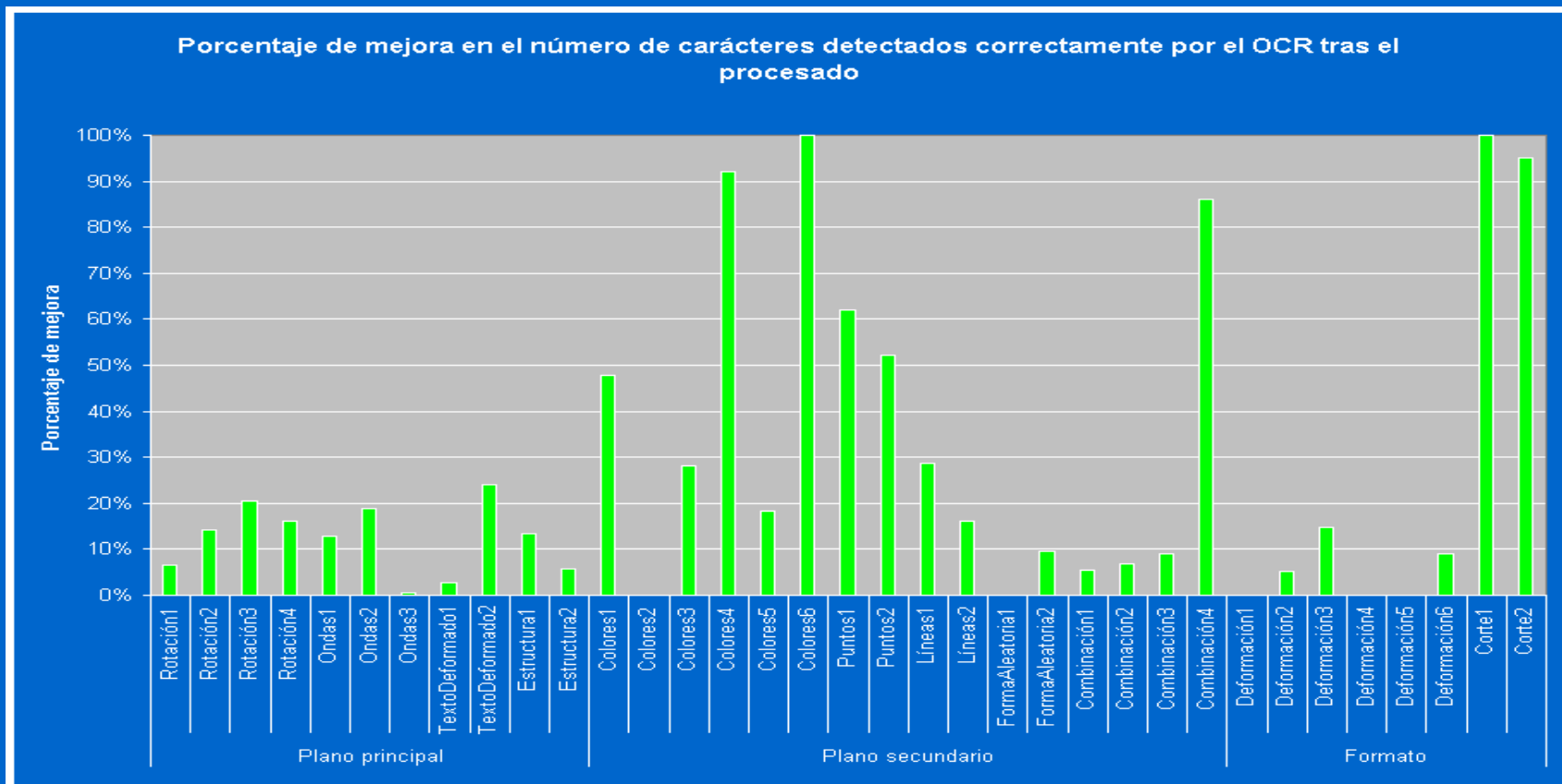
- Transformaciones en el formato:

- Corte:



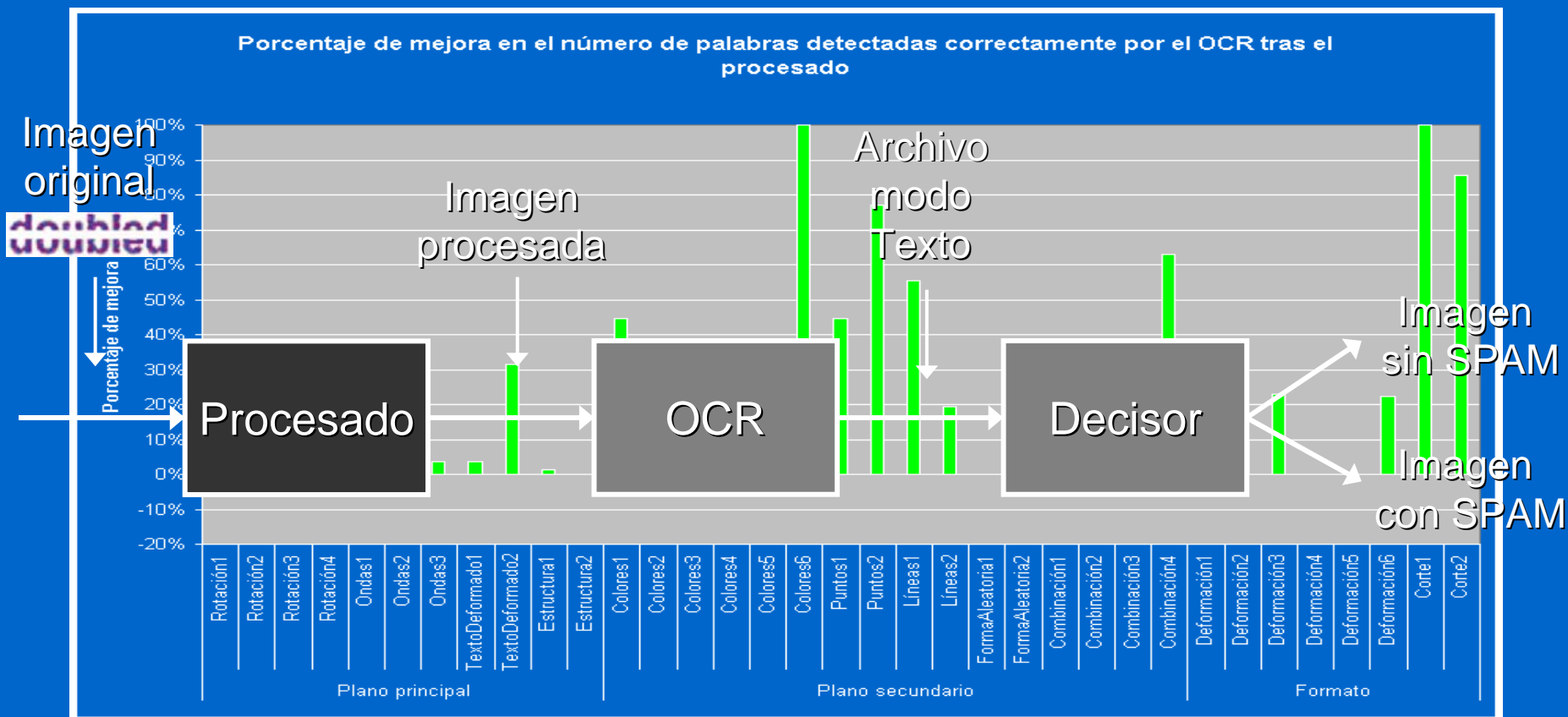
7. Resultados (1/3)

- Respecto al nº de caracteres:



7. Resultados (2/3)

- Respecto al nº de palabras:



7. Resultados (3/3)

- Casos especiales: Empeoramiento del porcentaje de palabras detectadas correctamente tras los procesados.

- Respecto al nº de caracteres :



Imagen original

$$\text{Tasa de acierto de caracteres del OCR} = \frac{36}{109} = 30,28 \%$$



Imagen procesada

$$\text{Tasa de acierto de caracteres del OCR} = \frac{47}{109} = 43,12 \%$$

- Respecto al nº de palabras:

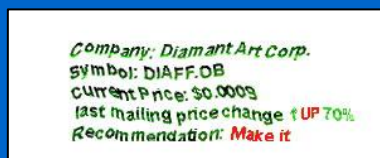


Imagen original

$$\text{Tasa de acierto de palabras del OCR} = \frac{3}{19} = 15,79 \%$$



Imagen procesada

$$\text{Tasa de acierto de palabras del OCR} = \frac{2}{19} = 10,53 \%$$

8. Conclusiones

- **Mejores resultados con las imágenes con modificaciones en el segundo plano.**
- **Calidad de la imagen interviene en los resultados.**
- **Una tasa alta de acierto con la imagen original, dificulta las mejoras.**



Dudas y preguntas





**Gracias por la
atención prestada**

