



# Optimización de la detección de bordes usando SSE

---

J. Enríquez - LU 36/08 - [juanenriquez@gmail.com](mailto:juanenriquez@gmail.com)  
N. Gleichgerrcht - LU 160/08 - [nicog89@gmail.com](mailto:nicog89@gmail.com)  
J. Luini - LU 106/08 - [jluini@gmail.com](mailto:jluini@gmail.com)

## Resumen

En la primera parte de este trabajo hicimos implementaciones de detección y graficación de bordes para ciertos operadores de derivación usando la arquitectura básica de la IA-32. Ahora el objetivo es optimizar dichas implementaciones (prácticamente hacerlas de nuevo) usando características más avanzadas de la familia IA-32: el modelo SIMD y el set de instrucciones SSE. Además, añadiremos a nuestro programa un operador de derivación nuevo que requiere procesamiento en punto flotante.

En este informe presentamos el trabajo realizado en esta segunda parte y las conclusiones obtenidas, incluyendo los problemas surgidos y la comparación de eficiencia entre estas nuevas versiones y las anteriores, entre las nuestras y las de la biblioteca OpenCV, y entre las implementaciones SSE de aritmética entera y la de punto flotante.

## Keywords

Detección de bordes - IA-32 - SIMD - SSE



**Facultad de Ciencias Exactas y Naturales**  
Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I)

Intendente Güiraldes 2160 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (54 11) 4576-3359

<http://exactas.uba.ar>

# Índice

<b>1. Introducción</b>	<b>2</b>
<b>2. Desarrollo</b>	<b>3</b>
2.1. Programa . . . . .	3
2.2. Implementación . . . . .	3
<b>3. Discusión</b>	<b>4</b>
<b>4. Conclusiones</b>	<b>6</b>
<b>5. Apéndice A: Manual de usuario</b>	<b>7</b>

## 1. Introducción

La detección de bordes en una imagen consiste en hallar las zonas de la misma en donde el color cambia “abruptamente”. Esta herramienta es utilizada tanto para compresión de archivos como para lograr efectos sobre las imágenes.

Una manera simple y muy útil de detectar bordes es calculando cómo varía la intensidad o luminosidad de una imagen entorno a cada uno de sus píxeles. Esto se logra tomando la imagen en escala de grises y aplicando en cada punto un operador de derivación, que es una matriz de números cuyo producto interno con el entorno del punto mide la variación de la intensidad en alguna dirección. Esta técnica se llama **convolución**.

Existen muchas matrices utilizadas para la convolución, con diferentes dimensiones y coeficientes. Las de mayores dimensiones proveen una detección más suave y menos precisa en la que se reducen los efectos del ruido”.

En el presente trabajo realizamos una implementación de los operadores de Roberts, Prewitt, Sobel y Frei-Chen en el lenguaje ensamblador correspondiente a la arquitectura IA-32. La misma contiene los algoritmos resueltos utilizando instrucciones de propósito general y cuenta además con las versiones de los mismos pero aplicando instrucciones específicas para procesamiento paralelo (SIMD).

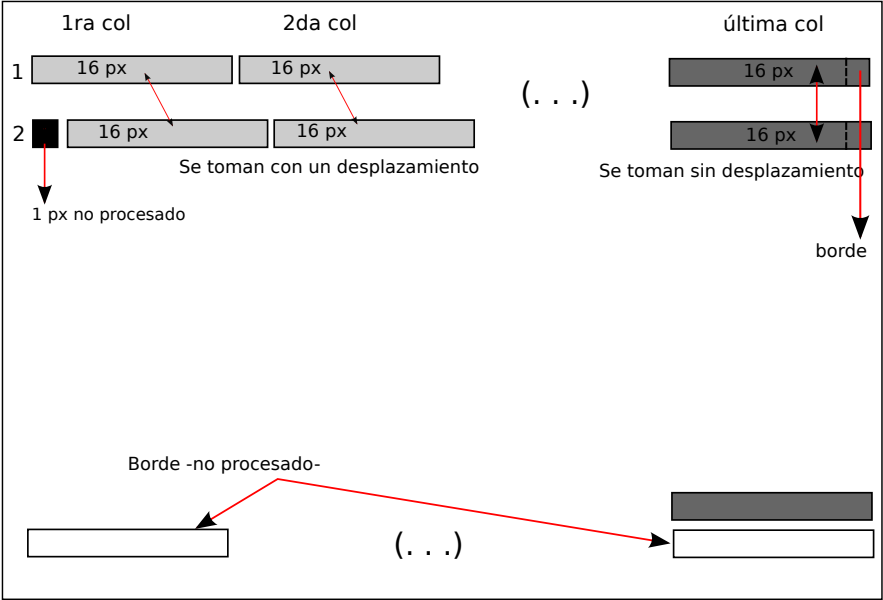
Respecto de la técnica SIMD (*Single Instruction, Multiple Data*), esta consiste en realizar operaciones a un vector de varios elementos de forma atómica, logrando esto mediante instrucciones y hardware especializado. De esta manera, se logra paralelizar a la hora de procesar información ya que como veremos más adelante, se puede llevar a cabo el cómputo de varios puntos de una imagen en un mismo paso, cosa que usando los registros de propósito general esto quedaba limitado a procesar de a uno por vez.

Como ventajas de este método nos gustaría destacar la capacidad del procesamiento paralelo que permite optimizar notablemente aplicaciones multimedia y de procesamiento de señales digitales. Si bien esto parece ser un gran y positivo cambio, cabe destacar que no todos los algoritmos son paralelizables en este sentido. Otro inconveniente importante es que el soporte para instrucciones SIMD en la arquitectura IA-32 se fue dando en forma progresiva ya que con el correr del tiempo se fueron agregando nuevas instrucciones. El problema que esto presenta es que al usar una determinada instrucción uno tiene que ver si ese procesador soporta esa instrucción en particular, lo cual, nuevamente, hace muy restrictivo y poco “portable” a los programas que utilizan SIMD.

En particular y para este trabajo, utilizamos instrucciones SIMD de la arquitectura IA-32 de Intel correspondientes a los subconjuntos de instrucciones denominados MMX, SSE1, SSE2 y SSE3. Consecuentemente, es requisito contar con un procesador de la familia Intel y similar que tenga soporte para estas instrucciones.

A continuación, describiremos brevemente el programa realizado, discutiremos las cuestiones surgidas durante su desarrollo y expondremos los resultados y conclusiones extraídos.

sadf



Ref:  
1. Fila actual  
2. Fila siguiente

asdf

## 2. Desarrollo

### 2.1. Programa

El programa que hicimos permite aplicar ciertos operadores de derivación a una imagen especificada por línea de comandos. Los operadores implementados son Roberts, Prewitt, Sobel y Frei-Chen. El primero tiene matrices de  $2 \times 2$  y detecta bordes en diagonal. Los otros tres tienen matrices de  $3 \times 3$  y detectan bordes verticales u horizontales. Se pueden invocar tanto las versiones de propósito general para los primeros tres, como las versiones de los operadores que utilizan instrucciones SIMD.

Para cada operador solicitado el programa aplica la matriz correspondiente en X, luego en Y, y finalmente suma los resultados. Para el caso de Sobel, también permite aplicar convolución solamente en X o solamente en Y.

Utilizamos la biblioteca **OpenCV** para manejar las imágenes convenientemente, es decir, dejamos la carga, el guardado de imágenes y el “aplanado” a escala de grises a dicha biblioteca. El grupo se limitó a crear una interfaz que contenga lo mencionado y a implementar los algoritmos de los operadores de detección de bordes en lenguaje ensamblador.

### 2.2. Implementación

### 3. Discusión

Tratándose de implementaciones en lenguaje en ensamblador, uno de las cuestiones que más tuvimos en cuenta y analizamos fue, por supuesto, la eficiencia de los algoritmos. Como ya dijimos, lo que hicimos fue medir la cantidad de clocks de procesador que cada función (incluso la de OpenCV y la que escribimos en C) consumen. Esto se logró mediante invocaciones desde C a la función `rdtsc` de lenguaje ensamblador.

El resultado que obtuvimos en ese sentido fue, como esperábamos, que nuestras implementaciones en ensamblador no alcanzaron la rapidez de la implementación de OpenCV, y que la implementación hecha en C fue aún más lenta. Los siguientes gráficos reflejan estas diferencias para imágenes de distintos tamaños.<sup>1</sup>

	Roberts	Prewitt	Sobel
<i>Media</i>	12618185	27276745	30886607
<i>Desvío</i>	1511854	2448652	2598392
<i>Max</i>	17743236	33694296	37954836
<i>Min</i>	11549460	25707576	28905552

	OpenCV	C	ASM
<i>Media</i>	26412756	65944312	30886607
<i>Desvío</i>	10563470	4921205	2598392
<i>Max</i>	68276028	80518500	37954836
<i>Min</i>	21301152	63370956	28905552

La diferencia de eficiencia entre la implementación en C y las hechas en ensamblador tiene un motivo claro y conocido. Por naturaleza, los lenguajes de programación de más alto nivel nos abstraen de muchas cuestiones de implementación, pero a la vez nos quitan control sobre esa implementación. Mientras en ensamblador usamos directamente los registros del procesador (memoria más rápida de la máquina) en C éstos se usan de manera interna y el programador trabaja con datos en memoria RAM.

Esta razón bastaría para justificar las diferencias de rendimiento, pero no es la única: además, el sólo hecho de estar programando en un lenguaje de nivel superior hace que el programador ponga menos atención en los detalles de eficiencia y más en otros aspectos como la legibilidad del código o su reutilizabilidad.

En cambio, el motivo de la diferencia de rendimiento entre nuestras implementaciones en ensamblador y la de OpenCV resulta menos evidente. De

<sup>1</sup> Las ejecuciones fueron realizadas en una PC: MSI Wind U100 con procesador Intel Atom N270 - 1Gb de RAM y bajo el S.O. Ubuntu Netbook Remix 9.04

hecho no estudiamos el código que utiliza OpenCV. Sabemos, sin embargo, que nuestras implementaciones no aprovechan al máximo las posibilidades de los procesadores, a diferencia de las de OpenCV que seguramente utilizan capacidades más avanzadas y/o específicas de los mismos como por ejemplo el uso de MMX, SSE, etc..

Un ejemplo que se nos ocurre es la saturación: nosotros, para cada píxel, verificamos si el resultado está por debajo de 0 o por encima de 255 para ajustarlo. Estos chequeos y ajustes consumen un tiempo significativo de ejecución; es claro que usando aritmética saturada nativa del procesador se mejoraría mucho la performance.

Otra característica que imaginamos se podría aprovechar es la paralelización: al tratarse de cálculos sencillos y repetitivos podrían venir bien las instrucciones de datos empaquetados.

También resulta interesante observar en qué medida se aceleró el algoritmo cuando eliminamos los accesos a memoria innecesarios que realizábamos al utilizar la pila en lugar de sólo los registros. El gráfico que sigue compara el rendimiento de la primera implementación en ensamblador que hicimos (que aplicaba Roberts en X usando la pila) con la definitiva (que minimiza los accesos a memoria y aplica Roberts en ambas direcciones). Es notorio que, aunque la primera implementación realiza menos cálculos pues aplica una sola matriz en lugar de dos, sea tanto menos eficiente por no aprovechar al máximo los registros del procesador.

	“Reg”	“Push”
<i>Media</i>	12618185	14692721
<i>Desvío</i>	1511854	1105330
<i>Max</i>	68276028	19269852
<i>Min</i>	11549460	13878600

## 4. Conclusiones

Como discutimos anteriormente, verificamos el hecho esperado de que nuestras implementaciones no alcanzarían la rapidez de las de OpenCV.

También es muy notorio cómo se puede llegar a mejorar la performance de un algoritmo implementándolo a tan bajo nivel. Para imágenes de tamaños típicos la ejecución realmente parece instantánea.

Para mejorar aún más las implementaciones creemos que el camino es aprovechar características más avanzadas de los procesadores modernos, en particular -y como ya mencionamos- la paralelización de cálculos y la utilización de una aritmética saturada más cercana al nivel de hardware. Suponemos que siguiendo este camino podríamos acercarnos mucho a la eficiencia lograda por OpenCV, aunque no tenemos aún las herramientas para verificarlo.



## 5. Apéndice A: Manual de usuario

El programa realizado permite aplicar diferentes implementaciones de detección de bordes a imágenes. Tanto el nombre de la imagen fuente como los operadores se indican por línea de comandos de la siguiente manera (suponiendo que el ejecutable está en `.exe/bordes`):

```
$ exe/bordes {fuente} {destino} {operadores}
```

Donde:

\* **fuente**: es la imagen de origen; puede tener cualquier formato soportado por la función `cvLoadImage` de OpenCV; los formatos más comunes están soportados;

\* **destino**: es el nombre de las imágenes de salida (generadas por el programa) SIN INCLUIR LA EXTENSIÓN; los archivos generados tendrán ese nombre más un sufijo que indica el operador utilizado, más la extensión; la extensión, el formato y el tamaño son obtenidos de la imagen de entrada;

\* **operadores**: una o más claves separadas por espacio; cada clave representa una implementación particular de detección de bordes; las claves posibles y sus significados son los que siguen:

CLAVE	OPERADOR	DIRECCION	IMPLEMENTACIÓN	SUFIJO
r1	Roberts	XY	Ensamblador	_asm_roberts
r2	Prewitt	XY	"	_asm_prewitt
r3	Sobel	X	"	_asm_sobelX
r4	Sobel	Y	"	_asm_sobelY
r5	Sobel	XY	"	_asm_sobelXY
cv3	Sobel	X	OpenCV	_cv_sobelX
cv4	Sobel	Y	"	_cv_sobelY
cv5	Sobel	XY	"	_cv_sobelXY
c3	Sobel	X	C	_c_sobelX
c4	Sobel	Y	"	_c_sobelY
c5	Sobel	XY	"	_c_sobelXY
push	Roberts	X	Ensamblador usando pila	_asm_roberts(push)
byn	Escala de grises			_byn

Por ejemplo, si tenemos el ejecutable en `"exe/bordes"` y una imagen en `"pics/lena.bmp"`, llamando al programa como

```
$ exe/bordes pics/lena.bmp pics/lena byn r1 r2 r3 r4 r5 cv3 cv4 cv5 c3 c4 c5 push
```

se aplicarán todas las implementaciones disponibles y se generarán los archivos `"pics/lena_asm_roberts.bmp"`, `"pics/lena_asm_prewitt.bmp"`, etc... Además el programa mostrará por salida estándar la cantidad aproximada de clocks de procesador insumida por cada implementación.