

Índice

1. Introducción	2
2. Smalltiles	3
2.1. Código C	3
2.2. Código ASM	3
2.3. Experimentación	3
2.3.1. Idea	3
2.3.2. Hipótesis	3
2.3.3. Resultados	3
3. Rotar canales	4
3.1. Código C	4
3.2. Código ASM	4
3.3. Experimentación	4
3.3.1. Idea	4
3.3.2. Hipótesis	4
3.3.3. Resultados	4
4. Pixelar	5
4.1. Código C	5
4.2. Código ASM	5
4.3. Experimentación	5
4.3.1. Idea	5
4.3.2. Hipótesis	5
4.3.3. Resultados	5
5. Combinar	6
5.1. Código C	7
5.2. Código ASM	7
5.3. Experimentación	9
5.3.1. Idea	9
5.3.2. Hipótesis	9
5.3.3. Resultados	9
6. Colorizar	10
6.1. Código C	10
6.2. Código ASM	10
6.3. Experimentación	10
6.3.1. Idea	10
6.3.2. Hipótesis	10
6.3.3. Resultados	10
6.3.4. Conclusión	11

1. Introducción

Este trabajo propone observar las diferencias en los tiempos de cómputo debidas al aprovechamiento del modelo de procesamiento SIMD con respecto a implementaciones en C.

Se implementaron diversos filtros de los cuales se expone una breve explicación a continuación y luego se detallan en sus respectivas secciones.

- **Smalltiles:** genera una imagen del mismo tamaño de la original que la contiene 4 veces, una en cada cuadrante.
- **Rotar canales:** intercambia los colores de modo que lo azul pasa a ser rojo, lo rojo, verde y lo verde, azul.
- **Pixelar:** como su nombre lo indica, pixela la imagen disminuyendo la definición.
- **Combinar:** consiste en combinar 2 imágenes en función de un parámetro.
- **Colorizar:** intensifica los colores predominantes en cada píxel.



Además de implementar los filtros en C y en Assembler se llevaron a cabo distintos experimentos para confirmar la hipótesis que motivó el trabajo, el hecho de que el procesamiento SIMD es más eficiente y permite resultados mucho más veloces que los obtenidos con las implementaciones de C.

2. Smalltiles

2.1. Código C

2.2. Código ASM

2.3. Experimentación

2.3.1. Idea

2.3.2. Hipótesis

2.3.3. Resultados

3. Rotar canales

3.1. Código C

3.2. Código ASM

3.3. Experimentación

3.3.1. Idea

3.3.2. Hipótesis

3.3.3. Resultados

4. Pixelar

4.1. Código C

4.2. Código ASM

4.3. Experimentación

4.3.1. Idea

4.3.2. Hipótesis

4.3.3. Resultados

5. Combinar

Este filtro consiste en realizar una combinación de dos imágenes que dependa de un número real α entre 0 y 255.

Cabe destacar que este filtro permite reutilizar cálculos debido a dos factores. Uno de ellos es la forma que tiene cada píxel generado en la imagen resultante, que consiste en que si se tienen una imagen A y una imagen B de tamaño $m \times n$ entonces el píxel de la imagen generada $I_{AB}^{i,j}$ se calcula de la siguiente forma:

$$I_{AB}^{i,j} = \frac{\alpha * (I_A^{i,j} - I_B^{i,j})}{255,0} + I_B^{i,j}$$

El otro factor es que nuestro filtro está optimizado para casos en los que la imagen B es el reflejo vertical de la imagen A . Es decir que se da que $I_A^{i,j} = I_B^{i,n-j+1}$ como se puede apreciar en la siguiente figura.

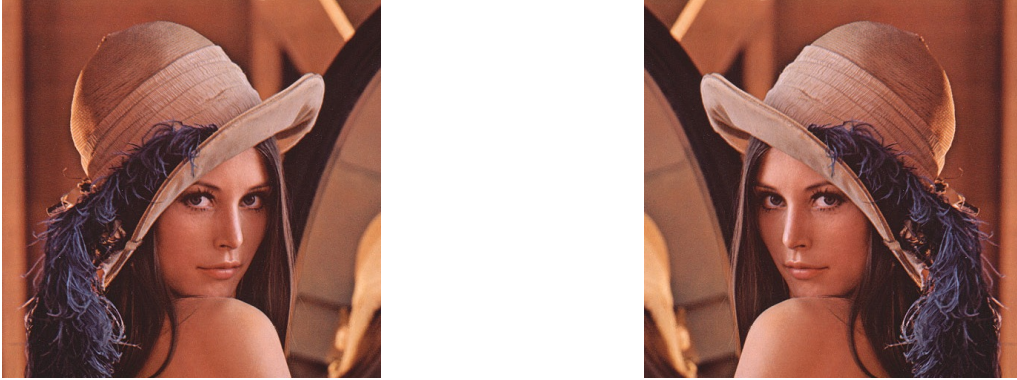


Figura 1: Se muestra la imagen A del lado izquierdo con la imagen B , el reflejo vertical de A , en la parte derecha.

Entonces se tiene que

$$\begin{aligned} I_{AB}^{i,j} &= \frac{\alpha * (I_A^{i,j} - I_B^{i,j})}{255,0} + I_B^{i,j} \text{ por la fórmula del filtro} \\ &= \frac{\alpha * (I_A^{i,j} - I_A^{i,n-j+1})}{255,0} + I_A^{i,n-j+1} \text{ dado que } I_A^{i,j} = I_B^{i,n-j+1} \end{aligned} \quad (1)$$

y análogamente

$$\begin{aligned} I_{AB}^{i,n-j+1} &= \frac{\alpha * (I_A^{i,n-j+1} - I_B^{i,n-j+1})}{255,0} + I_B^{i,n-j+1} \\ &= \frac{\alpha * (I_A^{i,n-j+1} - I_A^{i,j})}{255,0} + I_A^{i,j} \end{aligned} \quad (2)$$

Así, se obtiene

$$\begin{aligned} I_{AB}^{i,n-j+1} &= -1 * \frac{\alpha * [-1 * (I_A^{i,n-j+1} - I_A^{i,j})]}{255,0} - I_A^{i,n-j+1} + I_A^{i,n-j+1} + I_A^{i,j} \\ &= -1 * \left[\frac{\alpha * (I_A^{i,j} - I_A^{i,n-j+1})}{255,0} + I_A^{i,n-j+1} \right] + I_A^{i,n-j+1} + I_A^{i,j} \\ &= -1 * (I_{AB}^{i,j} - I_A^{i,n-j+1}) + I_A^{i,j} \text{ usando la igualdad de (2)} \end{aligned} \quad (3)$$

Estos cálculos muestran que luego de hacer el procesamiento para generar un píxel de la parte izquierda de la imagen resultante se puede obtener el píxel que corresponde a la mitad derecha con pocos cálculos más. Más aún, si se denomina

$$P = \frac{\alpha * (I_A^{i,j} - I_A^{i,n-j+1})}{255,0}$$

se consigue

$$I_{AB}^{i,j} = P + I_A^{i,n-j+1}$$

y

$$I_{AB}^{i,n-j+1} = -P + I_A^{i,j}$$

dando lugar a menos cálculos necesarios.

5.1. Código C

A continuación se presenta el pseudocódigo detallado del filtro.

Clamp es una función que controla la saturación.

Algorithm 1 Clamp

```

function CLAMP(pixel : float) → float
  if pixel < 0,0 then
    return 0,0
  else
    if pixel > 255,0 then
      return 255,0
    else
      return pixel

```

Combine realiza los cálculos correspondientes determinados por el filtro.

Algorithm 2 Combine

```

function COMBINE(a : unsigned char, b : unsigned char,  $\alpha$  : float) → float
  float af ← a
  float bf ← b
  return  $\frac{\alpha * (af - bf)}{255,0} + bf$ 

```

Combinar llama a las dos funciones ya explicadas y obtiene 2 píxeles con los que operar para luego dejar el resultado en la imagen destino.

Algorithm 3 Combinar

```

function COMBINAR(src: *unsigned char, dst: *unsigned char, cols: int, filas: int, srcRowSize: int, dstRowSize: int,  $\alpha$ : float)
  unsigned char (*srcMatrix)[srcRowSize] = (unsigned char(*)[srcRowSize]) src
  unsigned char (*dstMatrix)[dstRowSize] = (unsigned char(*)[dstRowSize]) dst
  for f ← 0 .. filas - 1 do
    for c ← 0 .. cols - 1 do
      bgrat * psa ← (bgrat*) & srcMatrix[f][c * 4]
      bgrat * psb ← (bgrat*) & srcMatrix[f][(cols - c - 1) * 4]
      bgrat * pd ← (bgrat*) & dstMatrix[f][c * 4]
      pd->b ← CLAMP(COMBINE(psa->b, psb->b,  $\alpha$ ))
      pd->g ← CLAMP(COMBINE(psa->g, psb->g,  $\alpha$ ))
      pd->r ← CLAMP(COMBINE(psa->r, psb->r,  $\alpha$ ))
      pd->a ← CLAMP(COMBINE(psa->a, psb->a,  $\alpha$ ))

```

5.2. Código ASM

Dado que cada píxel tiene 4 bytes y los registros XMM tienen 16 bytes, se levantan 4 píxeles contiguos desde la posición i, j y otros 4 píxeles desde la $i, n - j + 1$ en otro registro.

Se mostrará cómo se genera el píxel $I_{i,j}$ de la imagen resultante.

Se levantan 4 píxeles de la mitad izquierda en el registro XMM1:

XMM1:

A_{j+3}	R_{j+3}	G_{j+3}	B_{j+3}	A_{j+2}	R_{j+2}	G_{j+2}	B_{j+2}	A_{j+1}	R_{j+1}	G_{j+1}	B_{j+1}	A_j	R_j	G_j	B_j
16 0															

y 4 en espejo de la mitad derecha en XMM3:

XMM3:

A_{n-j+4}	R_{n-j+4}	G_{n-j+4}	B_{n-j+4}	A_{n-j+3}	R_{n-j+3}	G_{n-j+3}	B_{n-j+3}	A_{n-j+2}	R_{n-j+2}	G_{n-j+2}	B_{n-j+2}	A_{n-j+1}	R_{n-j+1}	G_{n-j+1}	B_{n-j+1}
16 0															

Teniendo

XMM9:

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16 0															

se ejecuta la instrucción **punpcklbw xmm1, xmm9** que da como resultado

XMM1:

0	A_{j+1}	0	R_{j+1}	0	G_{j+1}	0	B_{j+1}	0	A_j	0	R_j	0	G_j	0	B_j
16 0															

y luego **punpcklwd xmm1, xmm9** para seguir desempaquetando. De esta forma quedan 4 bytes para cada componente del píxel y se podrán convertir a floats para mayor precisión en los cálculos que implican el α .

XMM1:

0	0	0	A_j	0	0	0	R_j	0	0	0	G_j	0	0	0	B_j
16 0															

Se copió el contenido de XMM1 a XMM10 y con el registro que contenía al píxel $n - j + 4$, obtenido de forma análoga a como se obtuvo el píxel j

XMM8:

0	0	0	A_{n-j+4}	0	0	0	R_{n-j+4}	0	0	0	G_{n-j+4}	0	0	0	B_{n-j+4}
16 0															

se realizaron las restas entre componentes (**psubd xmm10, xmm8**) antes de convertir a float el registro XMM10 (**cvtdq2ps xmm10, xmm10**) dado que realizar la resta con float es una operación mucho más costosa que con enteros y además no había chances de perder precisión con la resta.

XMM10:

$A_j - A_{n-j+4}$	$R_j - R_{n-j+4}$	$G_j - G_{n-j+4}$	$B_j - B_{n-j+4}$
16 0			

Se multiplicó por α al registro XMM10 (**mulps xmm10, xmm0**) y luego se dividió por 255,0 (**divps xmm10, xmm14**) obteniéndose

XMM10:

$\frac{\alpha * (A_j - A_{n-j+4})}{255,0}$	$\frac{\alpha * (R_j - R_{n-j+4})}{255,0}$	$\frac{\alpha * (G_j - G_{n-j+4})}{255,0}$	$\frac{\alpha * (B_j - B_{n-j+4})}{255,0}$
16 0			

Para esto previamente, fuera del ciclo, se había movido el α que había llegado en los últimos 4 bytes de XMM0 a las otras 3 double words del registro con la instrucción **pshufd xmm0, xmm0, 00000000b** y se había puesto en XMM14 un 255.0 en cada double word declarando **mascara255: dd 255.0, 255.0, 255.0, 255.0** en **section .rodata** y luego haciendo **movdqu xmm14, [mascara255]**. De esta forma, al realizar estas operaciones fuera del ciclo, se minimizan los accesos a memoria y se evita repetir operaciones innecesarias.

Convirtiendo XMM8 a float con **cvtdq2ps xmm8, xmm8** y sumándolo a XMM10 con **addps xmm10, xmm8** se obtuvo

XMM10:

$\frac{\alpha * (A_j - A_{n-j+4})}{255,0} + A_{n-j+4}$	$\frac{\alpha * (R_j - R_{n-j+4})}{255,0} + R_{n-j+4}$	$\frac{\alpha * (G_j - G_{n-j+4})}{255,0} + G_{n-j+4}$	$\frac{\alpha * (B_j - B_{n-j+4})}{255,0} + B_{n-j+4}$
16			0

Antes de finalizar el procesamiento, se realizó **movups xmm15, xmm10** y **subps xmm15, xmm8** para conservar en XMM15 lo que en el desarrollo de cuentas apareció como P al explicar cómo se podían reutilizar cálculos.

Por último en el procesamiento de la parte izquierda se convirtió a entero XMM10 con **cvtps2dq xmm10, xmm10** y se empaquetó con los resultados de los demás píxeles usando instrucciones como **packusdw xmm11, xmm10** y **packusdw xmm13, xmm12** para pasar de double word a word,

XMM11:

F_{A_j}	F_{R_j}	F_{G_j}	F_{B_j}	$F_{A_{j+1}}$	$F_{R_{j+1}}$	$F_{G_{j+1}}$	$F_{B_{j+1}}$
16							0

luego **packuswb xmm13, xmm11** para pasar de word a byte,

XMM11:

F_{A_j}	F_{R_j}	F_{G_j}	F_{B_j}	$F_{A_{j+1}}$	$F_{R_{j+1}}$	$F_{G_{j+1}}$	$F_{B_{j+1}}$	$F_{A_{j+2}}$	$F_{R_{j+2}}$	$F_{G_{j+2}}$	$F_{B_{j+2}}$	$F_{A_{j+3}}$	$F_{R_{j+3}}$	$F_{G_{j+3}}$	$F_{B_{j+3}}$	
16																0

y finalmente **pshufd xmm11, xmm13, 0x1b** quedando todo listo para mover el contenido del registro a donde correspondiera.

XMM11:

$F_{A_{j+3}}$	$F_{R_{j+3}}$	$F_{G_{j+3}}$	$F_{B_{j+3}}$	$F_{A_{j+2}}$	$F_{R_{j+2}}$	$F_{G_{j+2}}$	$F_{B_{j+2}}$	$F_{A_{j+1}}$	$F_{R_{j+1}}$	$F_{G_{j+1}}$	$F_{B_{j+1}}$	F_{A_j}	F_{R_j}	F_{G_j}	F_{B_j}	
16																0

Para finalizar la parte derecha se multiplicaron todos los valores de las double words de XMM15 por -1 (**mulps xmm15, [menos1]** estando **menos1** definido en **rodata** como **menos1: dd -1.0, -1.0, -1.0, -1.0**) y se prosiguió con las conversiones y sumas como anteriormente.

5.3. Experimentación

5.3.1. Idea

5.3.2. Hipótesis

5.3.3. Resultados

6. Colorizar

6.1. Código C

El código C se trata de una conjunción de Fors, el exterior que recorre desde la segunda fila hasta la ante ultima, y el interior que recorre desde la segunda columna hasta la ultima, dejando así afuera a todos los bordes, tal como el enunciado decía. Luego en cada iteración del ciclo interior, que es donde se hacen las operaciones que modifican la imagen, lo que hacemos es crear un arreglo de unsigned chars, res", que es donde guardamos los máximos de cada canal en comparación a todos los píxeles lindantes del píxel en el cual estemos parado.

- Res [0] ← MaximoLindantesAzul
- Res [1] ← MaximoLindantesVerde
- Res [2] ← MaximoLindantesRojo

Luego con estos tres valores calculamos el alpha correspondiente de cada canal por el cual voy a multiplicar a cada uno. Y por último reescribimos el píxel final, en la imagen source con cada canal multiplicado por dicho alpha.

6.2. Código ASM

El código en ASM se trata también de una conjunción de ciclos. El recorre las filas desde la segunda hasta la anteÚltima, y el interior recorre las columnas desde la segunda hasta la anteúltima, pero saltando de a dos píxeles, que es la cantidad que procesamos simultáneamente con instrucciones SSE. El ciclo interior consta de tres partes, la primera es tan pronto se levanta de memoria los píxeles a procesar y todos sus lindantes, calculamos en dos registros los máximos de cada canal, de ambos píxeles a procesar con respecto a todos sus lindantes, y los guardamos ambos en un registro xmm. Luego la segunda parte es calcular el máximo de los máximos de ambos al mismo tiempo. Luego atraves de una par de operaciones, muy específicas para explicar, logro tener un registro xmm de dw, con cada dw representando un canal de cada píxel, en el orden establecido (argb), donde tengo un 1-alpha en la posición del canal que no tiene el máximo de los máximos, y un 1+alpha en la posición que tiene al máximo de los máximos, y luego concluyo multiplicando a cada píxel por su registro con los alphas indicados, lo reduzco a tamaño de 32 bits por píxel, los uno y los escribo en el destino.

6.3. Experimentación

6.3.1. Idea

En la experimentación de este filtro al igual que en el resto vamos a comparar el rendimiento respecto a los ciclos de clock, que tiene la función colorizar en C desde -o0 a -o3 contra asm. Sin embargo luego vamos a contrastar la función de asm, contra sí mismo pero primero agregando jumps de forma que no influya el flujo del programa solo para molestar al jump predictor. Luego vamos a correrlo tal cual está, y de a poco vamos a ir desenrollando el código. Como dijimos tiene un ciclo externo y uno interno, por lo que vamos a desenrollar primero el interno 2 veces, luego 4,16,32 y ver que pasa, y finalmente vamos a saltar a desenrollarlo completamente, cosa de no dejar casi ninguno controlador de flujo.

6.3.2. Hipótesis

Nuestra hipótesis es que el rendimiento va a ir mejorando a medida que vayamos cambiando los programas respectivamente a como los fui mencionando, o sea el más lento va a ser el código de asm, molestando al jmp predictor, y el más rápido el desenrollando el código 32 veces. El último test, por más que desenrollemos todo el programa creemos justamente que va a ser el más lento, por el tamaño del código, creemos que al hacer un código de tamaño mayor al de la cache, en un momento el pc va a estar haciendo muchísimos más miss en la cache, que los que ahorra sacando los controladores de flujo.

6.3.3. Resultados

Graficos lindos de lucia :D ciclos promedio colorizar c -o0 100 iteraciones : 101.898.160 ciclos promedio colorizar c -o3 100 iteraciones : 21.567.776 ciclos promedio colorizar original 1000 iteraciones: 5.980.263 ciclos promedio colorizar JodiendoJumps1 1000 iteraciones: 5.977.453 ciclos promedio colorizar Unrolling Ocho ciclos 100 iteraciones : 5.873.712

ciclos promedio colorizar Unrolling lena 1024x768 todos los ciclos 100 iteraciones : 35.285.958
ciclos promedio colorizar original lena 1024x768 100 iteraciones: 21.050.994

6.3.4. Conclusión

Conclusion Todo es una mierda. Conclusion posta: Primera gran conclusión reveladora, c -o0 demuestra ser mucho mas lento que optimizando, el código mejoro 5 veces su cantidad de ciclos por llamada al compilar con -O3. Luego sin embargo la función hecha en asm le sigue ganando por bastante a C, en una proporción de 3.5 veces mas rápido en cantidad de ciclos todavía. Ahora encunto al experimento personalizado, hicimos una comparación con distintos códigos para medir la influencia del jump predictor en la ejecución del código. Primera prueba fue ejecutar el mismo código que teníamos con el agregado de que en el ciclo interior (el que se repite mas veces) metíamos una serie de saltos transparentes con diferentes comparaciones, para tratar de romper con pipeline del procesador. Sin embargo por mas duro que itnentasemos descubrimos que el algoritmo que dirige al jump predictor es bastante mas sofisticado de lo que pensábamos, por lo que no pudimos influenciar en lo mas mínimo la eficiencia del código. Luego para contrastar el ultimo experimento mencionado, intentamos implementar lo contrario, desenrollar el código de manera que nunca influya un miss hit del jump predictor, en la ejecución del programa, sin embargo ya sacando la conclusion de lo mencionado recién, de que el sistema que lo regula es mas sofisticado del esperado, el desenrollar el código principal 2,4,8 veces, no estuvo mostrando ninguna diferencia con respecto al original. Por ultimo pasamos a otro experimento con desenrollar, en el cual se pretendia desenrollar completamente el codigo, generando un resultado en caunto a la cantidad de ciclos consumidos por llamada mucho mayor, por el tema de que un codigo tan grande no entraria completametne en la cache por lo que a partir de un punto el PC estaria generando constatnes miss hits en la cache y contrarestrando lo logrado evitar por el jump predictos. Y De hecho este experimento funciono muy bien, lo intentamos con una imagen mayor de lena, de 1024x768 para que el ciclo pueda ser desenrrollado mas cantidad de veces, y asi se ve que el codigo modificado tardo en promedio al rededor de unos 14.000.00 de ciclos mas que el original.