

ORGANIZACIÓN DEL COMPUTADOR II

*Departamento de Computación,
Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires*

TRABAJO PRÁCTICO 1: “OPORTUNCRISIS” (SEGUNDA ENTREGA)

Primer Cuatrimestre de 2009

Grupo "UNPCKHPD"

Pablo Antonio	290/08	pabloa@gmail.com
Pablo Herrero	332/07	pabloherrero@gmail.com
Estefanía Porta	451/04	estef.barbara@gmail.com

Índice

1. Introducción	1
1.1. ¿Qué se muestra ahora por pantalla?	1
1.2. El efecto <i>negative</i>	1
1.3. El efecto <i>smooth</i>	2
2. Desarrollo	2
2.1. Funciones desarrolladas	2
2.1.1. Función negative	2
2.1.2. Función smooth	2
2.2. Optimización	3
2.2.1. Realizando operaciones en paralelo (SIMD)	3
2.2.2. Evitando saltos	3
2.2.3. Acceder a memoria no es <i>tan</i> lento (memoria cache)	3
2.2.4. Operaciones aritméticas con lea y <i>shifts</i>	3
3. Resultados	4
4. Conclusiones	4

1. Introducción

La segunda entrega de este trabajo práctico consiste, básicamente, en:

- la *optimización* de las funciones que formaron parte de la primera entrega, y
- la inclusión de *dos nuevas funciones*: `smooth()` y `negative()`

La intención principal de este segundo trabajo es sacarle provecho a las distintas facilidades que proveen las extensiones SIMD¹ de la arquitectura x86 de Intel, así como aplicar distintas técnicas para la optimización del código que se ejecuta en dicha arquitectura.

1.1. ¿Qué se muestra ahora por pantalla?

Recordemos cuál era el ciclo que se repetía todo el tiempo en el juego:

```
1 procesar eventos de entrada
2 actualizar posiciones, estados, etc.
3 chequear IA, colisiones, física, etc.
4 mostrar resultados por pantalla
5 ira a 1.
```

Ahora, en la etapa 4 (“mostrar resultados por pantalla”), tiene lugar la siguiente secuencia (extendida):

```
1 generar el fondo actual (generarFondo)
2 generar el plasma, usando como color-off el color del cielo
  del fondo (generarPlasma)
3 para cada sprite de cada personaje:
4     recortar la instancia que se quiere del personaje
5     aplicar blit (cambiar el color-off del personaje por lo
      que haya en la pantalla)
6 si la tecla 1 o 3 se encuentra oprimida, comenzar la
  transición a un nuevo escenario
7 si la tecla 1 o 3 se encuentra oprimida, aplicar el efecto
  negative
8 si se esta realizando la transición a un nuevo escenario:
9     aplicar el efecto smooth
10    si hay más pixels negros que de otro color:
11        finalizar el cambio de escenario
```

1.2. El efecto *negative*

El resultado de aplicar la función `negative` sobre la pantalla guarda similitud con la imagen de los “negativos” de una fotografía (de las tomadas con cámara a rollo).

¹*Single Instruction, Multiple Data*

1.3. El efecto *smooth*

El efecto de la función `smooth` es *suavizar* la imagen. Puede pensarse en el efecto del suavizado como el contrario del “pixelado”; en una imagen suavizada, los píxeles no se diferencian con claridad.

2. Desarrollo

2.1. Funciones desarrolladas

2.1.1. Función `negative`

```
void negative();
```

Parámetros:

- Ninguno.

Archivo en el que se halla la función: `src/asm/negative.asm`

Pseudocódigo:

```
void negative():  
    para cada componente RGB c de cada pixel de la pantalla:  
        sumatoria = superior(c) + inferior(c) + anterior(c) + posterior(c) + 1  
        c = f(sumatoria)
```

Donde la función f es:

$$f(x) = \frac{1}{\sqrt{\text{sumatoria}}} \times 255$$

2.1.2. Función `smooth`

```
bool smooth();
```

Parámetros:

- Ninguno.

Archivo en el que se halla la función: `src/asm/smooth.asm`

Pseudocódigo:

```
bool smooth():  
    para cada componente RGB c de cada pixel de la pantalla:  
        sumatoria = superior(c) + inferior(c) + anterior(c) + posterior(c)  
        c = sumatoria/4  
        si sumatoria == 0:  
            contador_negros += 1  
        si no:  
            contador_blancos +=1  
    retornar evaluar(contador_negros > contador_blancos)
```

2.2. Optimización

2.2.1. Realizando operaciones en paralelo (SIMD)

Desde la aparición del procesador *Pentium II* de *Intel*, se agregaron a la arquitectura varias extensiones que proveen soporte para instrucciones SIMD. Entre ellas se encuentran: MMX, SSE, SSE2, SSE3 y SSE4.

Las instrucciones SIMD se utilizan para realizar operaciones en paralelo; se suele trabajar con vectores de datos en lugar de datos individuales. Las mejoras en *performance* son generalmente apreciables, y muchas veces determinan la utilidad de ciertas aplicaciones.

En general, las más beneficiadas por la utilización de instrucciones SIMD son las aplicaciones que realizan procesamientos similares de grandes cantidades de datos (por ejemplo, las aplicaciones multimedia).

En nuestro caso, las instrucciones SIMD nos permitieron procesar, en ocasiones, hasta 5 pixels en paralelo. Esto produjo la agilización de muchas de nuestras rutinas que trabajan sobre un gran número de pixels de la pantalla.

2.2.2. Evitando saltos

En la realización de este segundo trabajo, se trató especialmente de realizar la menor cantidad de saltos (incluidas las llamadas a subrutinas) posibles.

Sucede que la máquina realiza *fetches* de instrucciones que presume serán utilizadas en el corto plazo. Si bien existe un predictor de saltos, su heurística puede fallar. En esos casos, los saltos pueden llegar a interrumpir la secuencia de *prefetch* que realiza la máquina, haciendo que esta última pierda *performance*.

2.2.3. Acceder a memoria no es *tan* lento (memoria cache)

Hoy en día, las *memorias cache* son cada vez más grandes y rápidas. Si bien los accesos a la memoria principal suelen tomar varios ciclos de reloj, la realidad es que (siguiendo los *principios de localidad espacial y temporal*) las probabilidades de que los datos que buscamos se encuentren en alguna de las memorias cache de nuestra computadora, son altas.

Esto hace que el uso de memoria para almacenar datos temporales, útiles durante la ejecución de nuestro programa, cuando ya no disponemos de registros, no sea algo demasiado prohibitivo.

En lo que refiere a este trabajo, en ocasiones optamos por alojar datos (cálculos, información auxiliar) en memoria, sabiendo que, dependiendo de la operación, acceder a ellos podía ser más rápido que recalcularlos.

2.2.4. Operaciones aritméticas con *lea* y *shifts*

En ocasiones, es preferible usar *lea* para realizar operaciones aritméticas (en lugar de otras instrucciones específicas), en especial cuando el objetivo es conseguir una dirección

de memoria a partir de la cual obtener un dato, no sólo para ganar claridad, sino también porque puede ser más rápido al momento de la ejecución de nuestro programa.

También es importante tener en cuenta que, para multiplicaciones y divisiones por múltiplos de 2, es preferible utilizar *shifts* (aritméticos o lógicos, dependiendo el caso) en lugar de usar las instrucciones de multiplicación y división que son notablemente más costosas.

A lo largo del trabajo, tuvimos en cuentas ambas recomendaciones.

3. Resultados

Luego de realizar las optimizaciones antes mencionadas, se observan mejoras en la performance del juego.

La función **recortar**, encargada de proveer cada una de las imágenes de los *sprites* de los personajes, se ejecuta más rápidamente. Consecuentemente se observa mayor fluidez en el movimiento de los personajes.

Las funciones que no realizan ningún procesamiento de los pixels, sólo se ven beneficiadas por la reducción de saltos y los movimientos de memoria en tamaños mayores. En estos casos, la diferencia de *performance* es menos apreciable.

Las demás funciones, además de beneficiarse por las razones anteriores, agilizaron el procesamiento de los pixels. Como consecuencia, dado que las funciones son llamadas al menos una vez en cada ciclo de la aplicación, no se observan mayores retrasos en el refresco de la pantalla, como sí ocurría con las implementaciones que no sacaban provecho de las instrucciones SIMD.

En las mediciones, utilizando **rdtsc**, se observaron los siguientes resultados para la tardanza de las funciones²:

Función:	generarFondo	blit	recortar
Versión de la primera entrega:	8215504	444730	345994
Versión de la segunda entrega:	1967460	400664	212960
Diferencia:	6248044	44066	133034

4. Conclusiones

Estas son las conclusiones a las que arribamos luego de finalizar el trabajo:

- Las instrucciones SIMD son herramientas muy potentes que pueden proveernos *notables mejoras en la performance* de nuestras aplicaciones.
- Sin embargo, a veces, los cambios que son necesarios para sacarle provecho a dichas instrucciones *no son triviales*; la dificultad principal radica en la complejidad introducida por manipular numerosos datos en conjunto y el ingenio necesario para

²Se realizaron llamadas a las funciones para los mismos conjuntos de datos. Se utilizó la instrucción **rdtsc** en conjunt con **cuid** (esta última era necesaria para evitar la ejecución fuera de orden de la instrucción **rdtsc**).

aplicar las operaciones de la forma más adecuada para obtener los resultados deseados.

Referencias

- Intel 64 and IA-32 Architectures Software 1: Basic Architecture
- Intel 64 and IA-32 Architectures Software 2A: Instruction Set Reference, A-M
- Intel 64 and IA-32 Architectures Software 2B: Instruction Set Reference, N-Z
- Documentación del NASM: <http://www.nasm.us/doc/>
- http://en.wikipedia.org/wiki/Instruction_prefetch
- Intel 64 and IA-32 Architectures Optimization Reference Manual