

# Paralelización en Multiprocesadores de Dos Métodos para el Refinamiento de la Geometría en Reconstrucciones 3D

Luis Gerardo de la Fraga <sup>§</sup> y José María Carazo <sup>‡</sup>

<sup>§</sup> Centro de Investigaciones en Ingeniería y Ciencias Aplicadas,  
Universidad Autónoma de Morelos. Ave. Universidad 1001, 62210 Cuernavaca Mor.

E-mail: fraga@inaoep.mx

<sup>‡</sup>Unidad de BioComputación,  
Centro Nacional de Biotecnología-CSIC,  
Campus Universidad Autónoma, 28049 Madrid, España.

E-mail: carazo@cnb.uam.es

## Resumen

El refinamiento de la geometría de las proyecciones bidimensionales involucradas en el proceso de reconstrucción 3D, aplicada en este caso a la resolución de la estructura 3D de macromoléculas biológicas, es un proceso que consume gran cantidad de recursos. En este trabajo se presenta la paralelización por el paradigma de paso de mensajes de dos métodos de refinamiento de la geometría: 1) por comparación directa de las proyecciones [1] y 2) vía transformadas de Radon [2], sobre una plataforma multiprocesadora. En la paralelización se usa es esquema “master/slave” con el estándar PVM. Se presentan resultados del rendimiento de la paralelización de ambos métodos sobre una plataforma Power Challenge de SGI con seis microprocesadores.

**Palabra clave:** Procesado de imagen, Reconstrucción tridimensional, Refinamiento de la geometría, Paralelización, Paso de mensajes, Multiprocesadores.

## 1 Introducción

Una proyección de un *volumen* (un objeto tridimensional) es una colección de las estimaciones de las integrales de línea a través del volumen, realizándose la integración en la dirección del grupo de líneas que son perpendiculares a un plano en particular. La reconstrucción tridimensional (3D) es el proceso de estimar el volumen a partir de un grupo de proyecciones. Para poder realizar este proceso de la reconstrucción 3D, es necesario tener la información de la localización de las líneas relativas a un sistema

de coordenadas fijo al volumen. Sí se tiene una imagen que representa una proyección, la localización de las líneas perpendiculares al plano de proyección, que forman los pixels individuales de la imagen, queda expresada por lo que se conoce como los tres ángulos de Euler y por la localización en el plano de las dos coordenadas de la proyección del origen del sistema de coordenadas del volumen. Este problema es solucionado automáticamente en las reconstrucciones 3D médicas [3], como es la tomografía. En las aplicaciones médicas todas las proyecciones con tomadas de un mismo volumen (el paciente) y el sistema de coordenadas del volumen puede asumirse que está fijo al instrumento que colecta los datos.

En las reconstrucciones 3D de macromoléculas biológicas a partir de imágenes del microscopio electrónico de transmisión, el método de recolección de los datos va degradando al espécimen, limitándonos a no poder producir múltiples proyecciones de una misma partícula. En vez de ello nosotros debemos asumir que todas las partículas son idénticas, y de esta manera podemos considerar a una proyección de una partícula en particular como una de las proyecciones del volumen a ser reconstruido. El problema aquí es que se está muy limitado en el control de la orientación de las partículas de forma que se deben de usar las imágenes de proyección en sí mismas para estimar los ángulos de Euler y las coordenadas mencionadas más arriba.

Cuando se tiene un espécimen que muestre lo que se conoce como “vistas preferentes”, esto es, su orientación con respecto a la rejilla soporte del microscopio electrónico no es arbitraria, sino que interaccionan de una (o de unas pocas) formas determinadas, que dan lugar a una (o a unas pocas) “vistas preferentes”, se puede aprovechar esta característica para resolver su estructura 3D por medio de la técnica de reconstrucción basada en una geometría de inclinación cónica [4]. La estrategia se basa en tomar dos micrografías (fotografías provenientes microscopio electrónico), primero una inclinada un ángulo  $\phi$  y la segunda sin inclinar ( $\varphi = 0^\circ$ ). La primera micrografía proporciona las distintas vistas de proyección del espécimen que serán utilizadas en la reconstrucción 3D, La segunda micrografía, sin inclinar, proporciona los ángulos “rotacionales” mediante técnicas de alineamiento por correlación cruzada [5]. Los ángulos rotacionales forman el primer ángulo de Euler mientras  $\varphi$  es el segundo ángulo de Euler, común a todas las proyecciones. Las coordenadas del desplazamiento de las proyecciones inclinadas se obtiene por medio de la técnica del *estiramiento coseno*: la imagen inclinada es corregida por un factor  $1/\cos(\varphi)$  y correlacionada con la correspondiente imagen, ya alienada, sin inclinar. El tercer ángulo de Euler se hace cero y corresponde a la dirección del eje de inclinación de la micrografía.

Para el caso de que se tenga un espécimen que muestre en la preparación múltiples vistas “preferentes” se usa el método de *reconstitución angular* [6]. En este método los valores de los ángulos para cada proyección son asignados *a posteriori* basándose en las líneas comunes [7]. Las líneas comunes son una consecuencia directa del teorema de la sección central: dos proyecciones cualesquiera de un objeto representadas en el espacio de Fourier 3D se intersectan sobre una línea a través del origen, en el espacio recíproco, su *línea común*. Una vez que la línea común entre dos proyecciones es encontrada, se fija la rotación (en el plano) de dos secciones centrales de Fourier, quedando aún sin fijar el movimiento alrededor de la línea común. Introduciendo una tercera proyección y su sección central, que no sea paralela a ninguna de las otras

dos, fijará este movimiento y resultará en la completa determinación de los ángulos relativos entre las tres proyecciones (quedando solo una ambigüedad en cuanto a la mano del objeto). Partiendo de este sistema de orientaciones, nuevas proyecciones pueden adicionarse combinándolas con pares de proyecciones ya fijadas. Esto es, en esencia, el método de la reconstitución angular.

## 1.1 La necesidad del refinamiento

La metodología de adquisición de datos de la geometría de inclinación cónica nos produce una estimación menos que perfecta de los parámetros geométricos de las proyecciones con respecto al sistema de coordenadas fijas al volumen por varias razones: El ángulo de inclinación realmente no es idéntico para todas las partículas, sino que está dentro de un amplio margen que puede ser tanto como  $20^\circ$ , dado que el soporte no es completamente plano; los ángulos de rotación son determinados a partir de las imágenes sin inclinar, las cuales ya han sido previamente expuestas (mientras se tomó la imagen inclinada) al haz de electrones y pueden haber sido dañadas; la localización del origen de las proyecciones con respecto al sistema de coordenadas supuesto (ligado a las partículas) puede presentar serios problemas, algo que ciertamente ocurre al aplicar técnicas como el estiramiento coseno o los métodos descritos en [4, 8] a ciertos tipos de grandes macromoléculas. Adicionalmente, se usa un paso previo de clasificación de las partículas en base a su vista a  $0^\circ$ , un paso que puede dar resultados ambiguos, toda vez que en algunos casos los métodos de clasificación no consiguen realizar una separación precisa entre las diferentes vistas de un mismo objeto (ejemplos típicos serían imágenes muy ruidosas de criomicroscopía, por las pequeñas variaciones de la orientación en torno a la vista dominante, ó especímenes que generan imágenes de proyección parecidas a distintas orientaciones).

Aunque el método de reconstitución angular es elegante en su planteamiento conceptual, en la práctica no puede usarse para alinear imágenes individuales sin promediar, debido a la baja razón señal-a-ruido de estas imágenes. El método sí es aplicable a aquellas situaciones en que las imágenes a alinear tengan una relativamente alta relación señal-ruido, como sería el caso de partir de imágenes promediadas. Sin embargo, el promediado de imágenes puede implicar una pérdida de resolución al realizar el promediado entre imágenes con similares —pero no idénticas— orientaciones.

Todas estas limitaciones pueden ser resueltas en gran medida con el uso de un *método de refinamiento*. La esencia de tal método es el siguiente: tener un primer volumen estimado, calcular los parámetros geométricos necesarios por la consistencia de las proyecciones con las estimadas a partir del volumen, reconstruir con los nuevos parámetros y repetir hasta la convergencia. Dos de estos métodos se han desarrollado al mismo tiempo y ya se han utilizado con éxito a partir de imágenes individuales sin promediar: el método de comparación directa de las proyecciones [1] y el método de la transformada de Radon [2]. Con la adición de un método de refinamiento se puede lograr la resolución completa de los datos [9, p. 195].

## 2 Los Métodos de Refinamiento de la Geometría

Los métodos de refinamiento deben ser robustos, puesto que tienen que funcionar para imágenes típicas de criomicroscopía electrónica que presentan un SNR (Signal-to-Noise Ratio) menor que 1. Estos métodos están basados en la siguiente secuencia de la Fig. 1

<p><b>Paso 0:</b> Tener una reconstrucción tridimensional de referencia.</p> <p><b>Paso 1:</b> Proyectar la reconstrucción 3D de referencia en todas las direcciones de manera que cubran regularmente, equiespaciadamente, todo el espacio angular. Este grupo de proyecciones forman las <i>proyecciones de referencia</i>, o simplemente, <i>reproyecciones</i>.</p> <p><b>Paso 2:</b> Comparar cada proyección experimental (las que se usaron para tener la reconstrucción 3D del paso 0) con todas las reproyecciones. Los parámetros de la reproyección que más se ajusta, son los nuevos parámetros asignados a la proyección experimental dada.</p> <p><b>Paso 3:</b> Computar una nueva reconstrucción 3D con los nuevos parámetros asignados a las proyecciones experimentales.</p> <p><b>Paso 4:</b> ¿El 80% de las proyecciones experimentales cambia menos de 5° en el ángulo de inclinación? Si, PARAR. No, VOLVER al paso 1.</p>
--

Figura 1: Secuencia general de los métodos de refinamiento

Como una reconstrucción 3D de referencia puede usarse una estructura ya existente en baja resolución, el cual podría ser el resultado de la aplicación del esquema cónico aleatorio o de la reconstitución angular. También podría ser el resultado de unir varios grupos de proyecciones cónicas-aleatorias [10].

La orientación de una proyección queda expresada por los tres ángulos de Euler  $(\theta, \varphi, \psi)$ . Dos de estos ángulos,  $(\theta, \varphi)$ , son los responsables de la dirección de la proyección en el espacio 3D, mientras que el tercero  $(\psi)$  describe la rotación de la proyección “en el plano”. Por ende, basta con calcular las reproyecciones de la estructura 3D, en el paso 1 del método (ver Fig. 1), variando solo los dos primeros ángulos. Para una distribución equiespaciada de reproyecciones se escoge un incremento constante del ángulo  $\varphi$  y el incremento en el ángulo  $\theta$  es escogido de acuerdo a:

$$d\theta = \frac{d\varphi}{\sin \varphi} \quad (1)$$

La demostración de la Ec. (1) se encuentra en referencia [1]. Y para ser consistentes, el incremento del ángulo  $\psi$  se escoge igual al incremento del ángulo  $\varphi$ .

La forma de hacer robusto al método de refinamiento es utilizando la correlación cruzada para realizar la comparación entre las reproyecciones y las imágenes experimentales (de ahora en adelante llamadas simplemente como imágenes). La correlación cruzada es la mejor función para realizar la comparación, en el caso de tener datos contaminados con gran cantidad de ruido aditivo y gaussiano [5].

La condición de paro del método, paso 4 en la Fig. 1, la considera J. Frank [9, p. 243]. No se considera el cambio en los otros dos ángulos de Euler,  $\theta$  y  $\psi$  debido a que son responsables de la rotación de la imagen alrededor de un solo eje.

La reconstrucción tridimensional se lleva a cabo habitualmente en el espacio real utilizando la retroproyección convolucionada (weighted backprojection) [11], o técnicas algebraicas de reconstrucción [12].

El esquema de los métodos de refinamiento permite también la inclusión de nuevas imágenes sin la necesidad de conocer inicialmente su orientación, lo que hace posible el cálculo de una reconstrucción 3D de alta resolución para una macromolécula, de la manera siguiente: se usa una reconstrucción 3D ya calculada de la misma macromolécula como un “núcleo” y un nuevo grupo de imágenes colectado de micrografías únicas (y no de pares de micrografías como en el esquema cónico-aleatorio); tal “núcleo” es usado como una referencia en el primer paso 2 del método (ver Fig. 1).

## 2.1 El método basado en la comparación directa de las proyecciones

Este método combina dos procesos: (1) *El refinamiento de los parámetros de translación* y (2) El método de refinamiento en la orientación. El proceso (1) es bastante sencillo: la estructura de referencia es proyectada en la dirección de las proyecciones originales, y usando funciones de correlación cruzada, se corrigen los parámetros de translación. El proceso (2), más complicado, sigue los pasos de la secuencia mostrada en la Fig. 1. La comparación de las reproyecciones (paso 2, en la misma figura 1) con las imágenes experimentales es realizada así: Las reproyecciones y las imágenes a alinear son remuestreadas en coordenadas polares; se calculan las transformadas de Fourier de todas las proyecciones; la comparación se realiza por la función de correlación cruzada polar, lo que requiere una sola FFT inversa. La dirección de la imagen a alinear viene dada por el más grande coeficiente de correlación cruzada [nos resuelve los ángulos  $(\theta, \varphi)$ ]. El ángulo  $\psi$  está determinado por la posición del máximo en la correspondiente función de correlación cruzada polar.

En el caso de tener imágenes nuevas que se quieran agregar en la reconstrucción, se puede usar un método como el desarrollado en [13] para dar la primera estimación de los parámetros de translación.

El programa secuencial del método de refinamiento en la orientación sigue, de forma general, los pasos mostrados en la Fig. 2.

## 2.2 El método basado en la transformada de Radon

En principio este método es totalmente similar al método de comparación directa de las proyecciones descrito anteriormente. Aunque se distingue por su forma elegante de formular el problema y que le permite poner de manifiesto ciertas relaciones que no son evidentes en la otras aproximación.

El método de la transformada de Radon, [2] compara la transformada 2D de la proyección experimental con la transformada 2D de Radon de la proyección de referencia, extraída esta última de la transformada 3D de Radon del volumen de referencia. La

1. **Read** el volumen de referencia.
2. **Calculate** las reproyecciones del volumen de referencia equiespaciadamente en el espacio 3D. Se tienen  $I$  reproyecciones.
3. Se remuestran las reproyecciones a coordenadas polares y se realizan la FFT de los anillos.
4. **For each** imagen  $j$  a alinear:
  - (a) **Read** la imagen
  - (b) Se remuestran a coordenadas polares y se realiza la FFT de los anillos.
  - (c) *Calculate*:  $c_i = \text{CCF}(\text{reproyección}_i, \text{imagen}_j)$ ;
  - (d) **Find** el máximo coeficiente de correlación cruzada  $c_i$
  - (e) **Print** resultados.

Figura 2: Secuencia general del métodos de refinamiento en la orientación

comparación es realizada por correlación cruzada de dos transformadas 2D de Radon que permite hacer el alineamiento translacional (equivalente a “poner en fase” en el espacio de Fourier) simultáneamente con el alineamiento de la orientación.

La transformada 2D de Radon de una imagen es la versión discreta de la transformada general de Radon y es conocida también como “senograma” [6].

La teoría de este método viene desarrollada en [2], aquí se usará una descripción mucho más simple del método. Una de las propiedades de las transformadas de Radon (TR) es que las TR correspondientes a una proyección rotada por un ángulo adicional  $\alpha$  pueden ser obtenidas por permutaciones e inversiones de las líneas en la transformada extraída. Esta propiedad puede entenderse viendo el esquema de la Fig. 3, las líneas A,B,...,F, representan la transformada de Radon de una imagen muestreadas a  $\psi = 30^\circ$ , A está a  $0^\circ$ , B a  $30^\circ$  y así sucesivamente hasta F que está a  $150^\circ$ . La línea a  $180^\circ$ ,  $l(180^\circ)$ , es la reflejada a  $0^\circ$ , esto es, igual a  $\text{mirror}(l(0^\circ))$ . Si se calculara la correlación entre esta imagen, tomada como referencia, y otra imagen con índices A',B',... , F', se calcula la correlación cruzada entre las líneas AA', BB',... , FF'. De la misma forma, para calcular la correlación con la imagen de referencia rotada  $30^\circ$ , se tomaría la correlación cruzada entre las líneas BA', CB', DC', ED', FE' y  $\text{mirror}(A)F'$ ; y así sucesivamente. Si se realiza la búsqueda en la orientación, se promedian los coeficiente de orden cero de las correlaciones cruzadas de todas las líneas. Si se realiza la búsqueda del desplazamiento, se busca el máximo coeficiente promediado en un entorno dado en las correlaciones cruzadas de todas las líneas. Además, las propiedades de la transformada de Radon se mantienen para la transformada de Fourier (TF) de la TR, por lo que se usan las TF's de todas las TR para el cálculo de las correlaciones cruzadas.

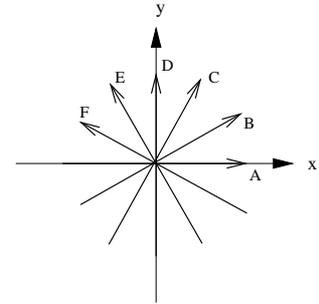


Figura 3: Representación esquemática de una transformada de Radon muestreada a  $\psi = 30^\circ$ .

La descripción general del método, en su forma secuencial, esta presentada en la Fig. 4

1. **Read** ref\_volume, la TF tridimensional de la TR tridimensional del volumen de referencia.
2.  $n_\psi$ , no. de valores en  $\psi = \frac{\psi_{\text{end}} - \psi_{\text{start}}}{\Delta \text{angles}} + 1$ ;
3.  $\Delta\varphi = \Delta \text{angles}$
4. **For each** imagen  $k$  a alinear:
  - (a) **read** image $_k$ ;
  - (b) **for**  $\varphi_{\text{start}} \leq \varphi_i \leq \varphi_{\text{end}}$ ,  $\varphi_{i+1} = \varphi_i + \Delta\varphi$ , **do**
    - i.  $\Delta\theta = \Delta\varphi / \sin |\varphi|$ ;
    - ii. **for**  $\theta_{\text{start}} \leq \theta_j \leq \theta_{\text{end}}$ ,  $\theta_{j+1} = \theta_j + \Delta\theta$ , **do**
      - A. reference\_img = do\_image ( ref\_volume,  $\theta_j$ ,  $\varphi_i$  );
      - B. **ccf\_radons**( reference\_img, image $_k$ ,  $n_\psi$  );
      - C. **find** el máximo de la función de correlación cruzada;
    - iii. **endfor**
  - (c) **endfor**
  - (d) **Print** resultados:  $(\theta, \varphi, \psi), (dx, dy)$ .

Figura 4: Secuencia general del método de refinamiento usando transformadas de Radon.

Como ha sido realizada, la secuencia del método contiene cuatro lazos principales. El primero (que comienza en el paso 4 en la Fig. 4) aplica la búsqueda de ángulos a cada imagen de entrada. Los otros tres lazos actúan sobre la búsqueda angular: primero sobre  $\varphi_i$  (en el paso 4.b), segundo en  $\theta_j$  (en 4(b).ii) y por último sobre  $\psi_n$  (dentro de la función **ccf\_radons** en 4(b).ii.B).

La función do\_image() (en 4(b).ii.A) extrae la imagen (una TF de la TR 2D) de referencia bidimensional, a los ángulos dados  $(\theta_j, \varphi_i)$ , de la TF 3D de la TR 3D de la estructura de referencia.

La función **ccf\_radons**() (que aparece en el paso 4(b).ii.B de la misma Fig. 4) realiza en sí misma la correlación cruzada entre las transformadas de Radon bidimensionales correspondientes a la proyección de referencia y la imagen a alinear.

## 3 Paralelización de los Métodos de Refinamiento

### 3.1 El método basado en la comparación directa de las proyecciones

Reportado en [1], el tiempo de cálculo para el *método de comparación directa de las proyecciones* viene a ser un poco más de dos días para 1000 proyecciones a alinear y para 5266 reproyecciones de referencia (que corresponde a un muestreo de  $2^\circ$  en el ángulo

$\varphi$ ). Estos corresponden a una estación de trabajo VAX 4000 modelo 60; que genera 1000 reproyecciones/hora y con un tiempo de 3 minutos para procesar la comparación de una imagen con 5266 reproyecciones, ya generadas, de referencia.

Aunque en [1] se menciona que el tiempo de cálculo es “aceptable”, éste calificativo hay que entenderlo en el contexto de la comparación con otro método de alineamiento que consumía 3 semanas para 1000 proyecciones [1]. De esta forma la idea de “paralelizar” los programas que conlleven más número de operaciones y/o consumo de recursos tiene una justificación clara.

Además hay que considerar la cantidad de almacenamiento necesario, ya sea en disco o en memoria. En la Tabla 1 se dan diversas cifras correspondientes al espacio necesario para diferentes intervalos de muestro del ángulo  $\varphi$ , suponiendo un tamaño de 16KB por reproyección (que corresponde aproximadamente a imagen  $64 \times 64$  pixels en punto flotante (4 bytes/pixel)). Este tamaño de imagen es normal para las aplicaciones con partículas aisladas, aunque hay campos de la microscopía electrónica que requieren imágenes mucho más grandes, de hasta  $512 \times 512$  pixels, para efectuar la reconstrucción 3D a una alta resolución.

$d\varphi$	No. de reproyecciones	Tamaño total (MB)
10.0	229	3.58
5.0	869	13.58
3.0	2365	36.95
2.5	3390	52.97
2.0	5266	82.28
1.5	9314	145.53
1.0	20847	325.73

Tabla 1: Volumen de datos que es necesario almacenar según el incremento en  $\varphi$  escogido

Con la estructura del método de la Fig. 2, este se paralelizó asignando equitativamente el trabajo a  $N_{\text{proc}}$  procesos “hijos” y un proceso “maestro” que coordina. Al principio el maestro lee y envía a los procesos hijos el volumen de referencia. Cada proceso hijo calcula una parte equitativa de las reproyecciones y realiza la comparación de la imagen a alinear con sus proyecciones locales. En esta parte, el proceso maestro solo manda a cada proceso hijo la imagen a alinear y recoge los resultados de la comparación, buscando en  $N_{\text{proc}}$  resultados el mayor coeficiente. La descripción general del método paralelizado puede verse en la Fig. 5

### 3.2 El método basado en la transformada de Radon

Una revisión del programa secuencial, a primera instancia, no da una guía obvia sobre como efectuar la paralelización. Nos encontramos, pues, ante un caso en que la estrategia de paralelización no es trivial y en el que es necesario realizar un análisis profundo de las dependencias entre las variables.

Desde el punto de vista del almacenamiento de datos, la cantidad de memoria requerida es relativamente pequeña: el volumen y dos imágenes bidimensionales. Si suponemos que tenemos un volumen de  $64 \times 64 \times 64$  voxels y las imágenes de  $64 \times 64$

### Proceso MAESTRO

1. **Read** el volumen de referencia.
2. **Send** el volumen de referencia a cada proceso hijo.
3. **Calculate** los ángulos  $(\theta, \varphi)$  y **send** a cada proceso hijo, equitativamente, ternas de valores  $(id, \theta, \varphi)$ .
4. **For each** imagen  $j$  a alinear:
  - (a) **Read** la imagen
  - (b) **Send** la imagen a cada proceso hijo.
  - (c) **Receive** de cada proceso hijo los resultados de la reproyección más parecida.
  - (d) **Find** la reproyección más parecida.
  - (e) **Print** resultados.

### Proceso HIJO

1. **Receive** el volumen de referencia.
2. **Receive**  $n$  ternas de valores  $(id, \theta, \varphi)$ .
3. **For**  $0 \leq i < n$ 
  - (a) **Calculate** la  $i$  reproyección
  - (b) Se remuestrea a coordenadas polares y se realiza la FFT de los anillos.
4. **endfor**
5. **While receive** una imagen a alinear:
  - (a) Se remuestrea la imagen a coordenadas polares y se realizan la FFT de los anillos.
  - (b) **Calculate**:  $c_n = \text{CCF}(\text{reproyección}_n, \text{imagen})$ ;
  - (c) **Find** el máximo coeficiente de correlación cruzada  $c_n$
  - (d) **Send** al proceso MAESTRO los resultados.

Figura 5: Descripción general del método de refinamiento en la orientación paralelizado

pixels, tomando 60 muestras en la transformada de Radon a un muestreo angular de  $2^\circ$ ; la cantidad de memoria requerida para almacenar una transformada de Radon será  $60 \times (180^\circ/2^\circ) \times 4 = 21,600$  bytes ó, aproximando, 21KB. La transformada de Fourier requerirá  $(60 + 2) \times (180^\circ/2^\circ) \times 4 = 22,320B \approx 22KB$ . Y la transformada de Radon del volumen requerirá  $(180^\circ/2^\circ) \times 21600B$ , esto es 1.85MB y su transformada de Fourier 1.92MB. En total, menos de 2MB en datos. Por tanto, no parece que el tema de almacenamiento de datos sea importante en este problema.

A continuación pasamos a estudiar los requerimientos de cálculo del proceso. Para ello realizamos un “profiling”, o “perfilado”, del programa secuencial. Los resultados del perfilado indicaron que el 99.18% del tiempo total de ejecución eran consumidos por la función `ccf_radons` (ver Fig. 4). El trabajo principal de esta función es la multiplicación compleja de los coeficientes de Fourier de las líneas de las transformadas de Radon 2D. También, dentro de esta misma función, se llama a otra función que realiza la transformada inversa de Fourier, para cada par de líneas, y la carga de ejecución es grande debido al número de veces que se realiza la FFT inversa. El algoritmo usado para realizar la FFT es multiradix, lo que no nos limita que el tamaño de las líneas sea una potencia de 2, pero tiene el inconveniente de un mayor tiempo de ejecución.

Así, con los resultados del perfilado, se realizó la paralelización explícita de la función `ccf_radons`. El proceso maestro consiste básicamente de la secuencia mostrada en la Fig. 4, excepto de lo siguiente: el proceso maestro envía a todos los procesos hijos la imagen extraída de la transformada de Radon 3D del volumen; y se crea la nueva función `ccf_radons_parallel`. La función `ccf_radons_parallel` en el proceso maestro se encarga de recibir los resultados parciales del procesamiento de cada proceso HIJO. La función `ccf_radons_parallel` en los procesos hijos es la que realiza la carga pesada: sí las TR's tienen  $n$  líneas, cada proceso hijo realiza las correlaciones cruzadas de  $n/N_{\text{proc}}$  líneas y envía los resultados al maestro. El estudio detallado de esta paralelización se encuentra en [14].

## 4 Resultados

Todos los programas, de forma secuencial y paralelos, se codificaron en lenguaje C. Como ambiente de paso de mensajes se utilizó PVM [15]. Los programas se ejecutaron en una máquina *Power Challenge*, de Silicon Graphics, con seis procesadores R10000.

En la Fig. 6(a) se visualiza la gráfica (log-log, como se sugiere en [16]) del tiempo de ejecución necesario para realizar 5266 reproyecciones versus el número de procesos hijos. Corresponde al trabajo realizado en el primer ciclo *for* en el pseudocódigo del proceso hijo (en la Fig. 5, 3). En la Fig. 6(b) se presenta el tiempo de ejecución, se presenta la media  $\pm$  la desviación estándar de varias mediciones ( $> 5$ ), de comparar 1000 proyecciones contra las 5266 reproyecciones de referencia, y que corresponde al ciclo *while* en el pseudocódigo del proceso hijo (mis Fig. 5, línea 5

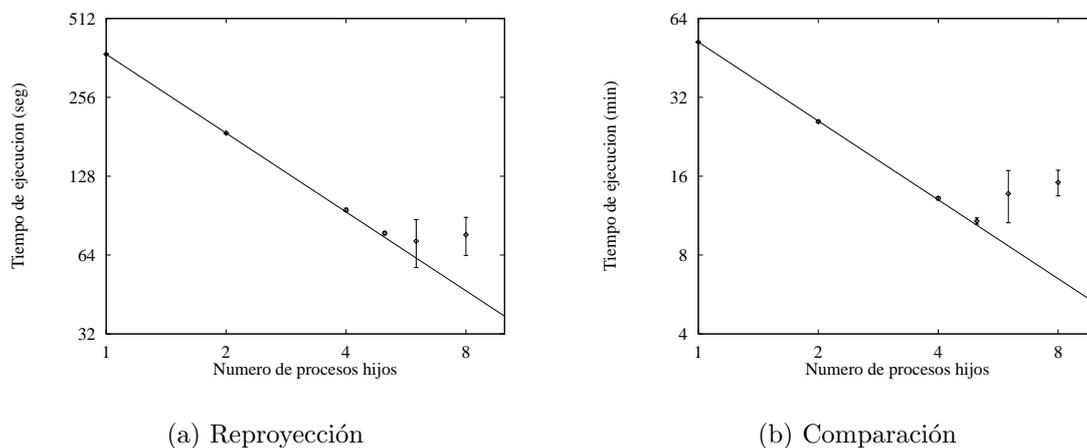


Figura 6: Tiempos de ejecución del *Método de Refinamiento de la Orientación*. Se presenta la media  $\pm$  la desviación estándar de varias mediciones ( $> 5$ ). La línea recta indica la aproximación ideal para la ejecución del programa paralelo.

La medición del tiempo de ejecución correspondiente a un proceso hijo de la Fig. 6 se realizó de dos maneras: (1) ejecutando el programa paralelo con un sólo proceso hijo y (2) con una versión secuencial del programa. Los tiempo de ejecución son semejantes,

sin diferencias apreciables, entre (1) y (2), lo que indica que la carga debida a las comunicaciones en el programa paralelo es mínima.

Los resultados del método vía transformadas de Radon se presentan en la 7. Se presenta la media y desviación estándar en el eje-y de varias mediciones del tiempo de ejecución. Se midió el tiempo total de ejecución (wall clock time), como se sugiere en [16], que involucra también los tiempos en Entrada/Salida, paginado, inicialización del sistema PVM, etc. Las condiciones para el experimento fueron: un volumen de  $64 \times 64 \times 64$  voxels y para *una sola* imagen de  $64 \times 64$  pixels, tomando 60 muestras en la transformada de Radon a un muestreo angular de  $2^\circ$ . Los parámetros de entrada al programa fueron:  $0^\circ$ — $90^\circ$  en  $\varphi$ ,  $0^\circ$ — $180^\circ$  en  $\theta$  y  $-90^\circ$ — $90^\circ$  en  $\psi$ . Se utilizó un muestreo angular de  $2^\circ$ .

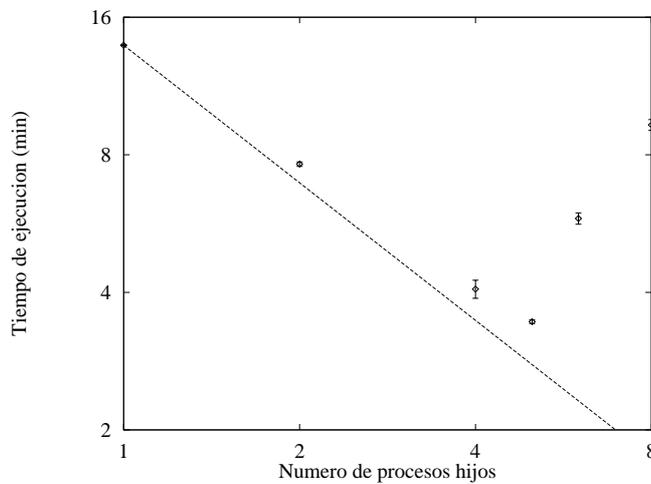


Figura 7: Resultados de la medición de tiempos de ejecución vs. el número de procesos hijos de los programas secuencial y paralelo para el método de refinamiento vía transformadas de Radon. La línea recta indica la aproximación ideal para la ejecución del programa paralelo.

Se presenta la media  $\pm$  la desviación estándar de varias mediciones ( $> 5$ ).

En la Fig. 7 se presenta la media y desviación estándar en el eje-y de varias mediciones del tiempo de ejecución. Se presenta el tiempo total de ejecución, como se sugiere en [16], que involucra también los tiempos en Entrada/Salida, paginado, inicialización del sistema PVM, etc. Las mediciones fueron realizadas bajo una condición de carga “baja” en la plataforma.

La caída en el rendimiento teniendo más de cinco procesos hijos que se observa en las Figs. 6 y 7 se explicara a continuación. En nuestro escenario tenemos: i) el proceso maestro, ii) los procesos hijos, iii) los procesos del demonio de PVM para el ambiente de paso de mensajes y iv) los demás procesos que incluyen los del sistema operativo y los de otros usuarios (el uso de la máquina es compartida por varios grupos de trabajo). La máxima eficiencia se logra con cinco procesos hijos más el proceso maestro, esto es, un proceso por procesador (recordemos que la plataforma tiene seis microprocesadores). Pero la carga del proceso maestro es mínima, solo hace la colecta de los resultados de los procesos hijos, por lo que aún tendríamos que tener un buen resultado con seis procesos hijos. Sin embargo hay que tener en cuenta la carga debido a iii) y especialmente iv).

Con iii) y iv) en nuestro escenario, la máxima eficiencia se logra con cinco procesos hijos, como queda demostrado en la gráficas. Todas las mediciones fueron realizadas bajo condiciones de carga “baja” en la plataforma (las mediciones fueron realizadas fuera del horario laborable).

Merece destacarse en este apartado que se estudió la posibilidad de incorporar a la aplicación paralela un control explícito del balanceo de la carga. Se pretendía de esta forma anticiparse a situaciones reales de uso de un máquina multiprocesador en un entorno de acceso y ejecución no exclusivo. Sin embargo, se observó que la propia gestión de los procesos que realiza el sistema operativo de la máquina (Irix 6.3) era bastante eficiente. En concreto, el sistema operativo “ve” todo el hardware como una sola máquina, esto es, los procesos no quedan residentes en un solo procesador, sino que la cola del despachador reasigna dinámicamente cada proceso al procesador disponible. Este hecho se verificó estudiando exhaustivamente el patrón de ejecución del programa paralelo de refinamiento en la orientación. Con menos de siete procesos hijos, se mantiene una rebanada del proceso por procesador; arriba de seis procesos hijos, los multiplexa en los procesadores. Así, como no se tiene control directo del usuario sobre los procesadores, lo que se buscó finalmente fue equidistribuir la carga. En las estrategias de paralelización tratadas, esta equidistribución se logró fácilmente haciendo que cada proceso hijo mantuviese un número igual de reproyecciones, para el caso del método de refinamiento de la orientación, y para el caso del método por transformada de Radon se logró asignando a cada proceso uno un número igual de líneas de la transformada de Radon 2D.

## 5 Conclusiones

Se han analizado dos métodos de refinamiento de la geometría de la estructura tridimensional de macromoléculas biológicas. Los dos métodos son los existentes actualmente: 1) El método basado en la comparación directa de las proyecciones, debido a Penczek *et al* [1], y 2) El método basado en las transformadas de Radon, debido a Radermacher [2].

Se han realizado los programas secuenciales y paralelos de ambos métodos. La paralelización se ha realizado bajo el paradigma de paso de mensajes utilizando las librerías de PVM; de forma que la aplicación puede ser fácilmente transportable a un amplio conjunto de máquinas paralelas. Por las mediciones realizadas se tiene un muy buen rendimiento de los programas paralelos programándolos por paso de mensajes en una máquina con arquitectura del espacio de direcciones compartido, una máquina multiprocesadora, como la Power Challenge de SGI.

## Referencias

- [1] P. A. Penczek, R. A. Grassuni, and J. Frank. The ribosome at improved resolution: new techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy*, 53:251–270, 1994.

- [2] M. Radermacher. Three-dimensional reconstruction from random projections: Orientational alignment via Radon transforms. *Ultramicroscopy*, 53:121–136, 1994.
- [3] G. T. Herman. *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*. Academic Press, 1980.
- [4] M. Radermacher. The three-dimensional reconstruction of single particles from random and non-random tilt series. *J. Electron Microsc. Tech.*, 9:359–394, 1988.
- [5] W. O. Saxton. *Computer Techniques for Image Processing in Electron Microscopy*. Academic Press, 1978.
- [6] M. van Heel. Angular reconstitution: a posteriori assignment of projection direction for 3D reconstruction. *Ultramicroscopy*, 21:111–124, 1987.
- [7] R.A. Crowther, L.A. Amos, J.T. Finch, D.J. De Rosier, and A. Klug. Three dimensional reconstruction of spherical viruses by fourier synthesis from electron micrographs. *Nature*, 226:421–425, 1970.
- [8] J. M. Carazo and J. Frank. Three-dimensional matching of macromolecular structures obtained from electron microscopy. *Ultramicroscopy*, 25:13–22, 1988.
- [9] J. Frank. *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*. Academic Press, 1996. San Diego, CA.
- [10] Felix de Hass, J.-C. Taveau, N. Boisset, O. Lambert, S.N. Vinogradov, and J. Lamy. Three-dimensional reconstruction of the chlorocruorin of the polychaete annelid *eudistylia vancouverii*. *J. Mol. Biol.*, 225:140–153, 1996.
- [11] M. Radermacher. Weighted back-projection methods. In J. Frank, editor, *Electron Tomography*, pages 91–115. Plenum, 1992.
- [12] R. Marabini, G.T. Herman, and J.M. Carazo. 3D reconstruction in electron microscopy using ART with smooth spherically symmetric volume elements (blobs). *Biophys. J.*, 1997. In Press.
- [13] L.G. de la Fraga, J.M. Carazo, and G.T. Herman. A method for obtaining a low resolution 3d reconstruction from very noisy projections. In *Proc. VII National Symposium on Pattern Recognition and Image Analysis*, volume 1, pages 347–352, 1997.
- [14] L.G. de la Fraga. *Refinamiento de la Estructura Tridimensional de Macromoléculas Biológicas Aisladas: Principios y Realización en Multiprocesadores*. PhD thesis, Universidad Autónoma de Madrid, Facultad de Informática, 1998.
- [15] V.S. Sunderam. Network-based concurrent computing on the pvm sytem. *Concurrency: Practice and Experience*, 2(4):293–311, 1992.
- [16] L.A. Crowl. How to measure, present, and compare parallel performance. *IEEE Parallel & Distributed Tech.*, 2:9–25, 1994.