

Tarea 2

ANOVA, 2^k

Fecha de entrega: jueves de semana 12, 18:00 horas.

Un científico de la computación, apasionado por el área de gráficos por computadora, se propone crear un algoritmo de procesamiento híbrido (hace uso del CPU y GPU integrado) para mejorar el desempeño en computadoras que no cuentan con GPU discreto. El algoritmo de sintetizado de imágenes por computadora que utilizó hace uso de las propiedades físicas de la luz para generar la imagen, por lo que clásicamente se han utilizado CPUs cuando se hace uso de este algoritmo, sin embargo, en investigaciones recientes se han utilizado GPUs modernos para sintetizar esta clase de imágenes.

Dado que el objetivo de su investigación es mejorar el rendimiento del sintetizado de imágenes en sistemas que únicamente cuentan con GPU integrado (conocidos en la industria como SoC o APUs), la variable de respuesta de su experimento es tiempo de sintetizado. Con la finalidad de hacer la comparación de la manera más justa posible, utilizó un CPU, APU y GPU del mismo rango de precios. El algoritmo híbrido corre únicamente en el APU.

Varios factores impactan el tiempo de sintetizado: la cantidad de objetos presentes en la escena (la escena en sí es definida por este factor), los efectos visuales de la escena: anti-aliasing (AA), reflexiones (RE) y transparencias (TR). De igual manera, decidió sintetizar la misma imagen en tres resoluciones diferentes: 1280x720, 1440x900 y 1920x1080. Por último, la misma escena, con los mismos escenarios se ejecutaron en las tres diferentes arquitecturas (CPU, GPU y APU).

El razonamiento del científico para incluir tantos factores en su experimento es que las imágenes sintetizadas por computadora son muy diversas por naturaleza, y quería estar completamente seguro de que su propuesta de algoritmo tenía la capacidad de brindar mejor rendimiento en distintos escenarios, y no convenientemente en uno específico.

El científico automatizó la ejecución de los experimentos, garantizando aleatoriedad, y que cada escenario específico corriera un número repetido de veces (de las cuales no se acuerda).

Peor aún, los datos que el científico recabó no están listos para ingresarse a R, sino que se necesita de un preprocesamiento.

Cada medición tomada cuenta de cuatro líneas, como las que se muestran a continuación:

APU image written, Rays: 2758193

real 6,53

user 50,25

sys 0,33

10167_16000APU010-1280x720.sc

Las líneas en rojo no sirven para el análisis que se desarrolla en este documento y se pueden eliminar. La línea que dice “real” contiene el tiempo de sintetizado de la imagen (nuestra variable de respuesta).

De la última línea, la información relevante es la que se encuentra después del _.

La nomenclatura es _(cantidad de objetos)(arquitectura)(efectos en la escena)-(resolución de la escena).

Los efectos de la escena tienen una codificación binaria, donde 1 significa presente y 0 no presente. El orden es (anti-aliasing)(transparencias)(reflexiones). 000 significa no hay ningún efecto y 111 que todos los efectos están presentes.

Preguntas que deben contestarse en el documento:

¿Es el tiempo de sintetizado lo único que se debe verificar en este experimento?

¿Cuántas repeticiones por escenario realizó el científico?

¿Cómo es el comportamiento de la normalidad y homocedasticidad en los datos?

¿Existe alguna transformación que cumplir los supuestos necesarios? De existir, aplíquela a los datos.

¿Es la propuesta del científico mejor en todos los escenarios? ¿En cuáles sí, en cuáles no? Brinde sustento estadístico para sus respuestas.

Usted como estudiante crítico es escéptico de la descripción del experimento del cual le están brindando los datos para su análisis.

- 1. ¿Qué considera que se está pasando por alto?*
- 2. ¿Considera que la comparación entre arquitecturas es justa?*
- 3. ¿Qué información falta con respecto a lo que corre en el CPU y GPU?*
- 4. ¿Por cuál otra razón usted rechazaría estos resultados?*

Requisitos indispensables:

El incumplimiento de alguno de los siguientes requisitos vuelve a la asignación como “no revisable” y se obtiene automáticamente una calificación de cero.

1. Todo el análisis debe realizarse utilizando R.
2. Se debe incorporar dentro del documento el código utilizado para obtener los resultados.
3. El documento no cuenta con los estándares esperados de un estudiante de maestría tales como orden, redacción clara y cuidado en cuenta a redacción y ortografía en su elaboración.

4. El documento debe ser redactado en Latex haciendo uso del formato IEEE transactions. Se deben proveer los archivos fuentes del documento.

Procedimiento:

1. En la carpeta de tareas del TEC digital para este curso hay un archivo llamado "datos tarea 2", este es el que se utilizará para esta tarea.
2. Procese los datos para poder brindarlos como entrada a R.
3. Analice el experimento y cree la tabla ANOVA. Identifique si hay algún factor que se modeló como 2^k .
4. Realice el análisis de los supuestos del ANOVA. Aplique transformación a los datos en caso de ser necesario.
5. Realice el análisis de ANOVA en R ¿Qué se puede concluir? ¿Qué información se obtiene? ¿Tienen sentido todas las comparaciones entre factores? ¿Cuáles comparaciones tiene sentido hacer, cuáles no?
6. Obtenga el gráfico de medias y bigotes del tiempo en función de la arquitectura y de los otros factores del experimento (un gráfico por cada otro factor), ¿intuitivamente qué se observa? ¿Se puede concluir algo?
7. Investigue las *pairwise t-tests* y aplíquelas entre los escenarios que tenga sentido hacerlo. Nuevamente, el tiempo en función de la arquitectura y de los otros factores (una tabla por comparación de los otros factores) ¿Qué se puede concluir?
8. Obtenga el gráfico de medias y bigotes del tiempo, en función de únicamente la arquitectura ¿Se podría utilizar únicamente este gráfico para concluir algo?
9. Obtenga métricas (propuestas por usted) del desempeño de la arquitectura APU en comparación con las otras dos.
10. Analice y concluya sobre los resultados obtenidos.

Evaluación:

- a. Cumplimiento del procedimiento: 30%.
- b. Gráficos de buena apariencia, presentados con un formato que facilita su visualización e interpretación: 40%.
- c. Correcta aplicación de la teoría estadística: 10%
- d. Conclusiones y análisis: 20%

El pdf se sube al Tec Digital, los archivos fuentes se envían en una carpeta comprimida al correo del profesor errivera@itcr.ac.cr