

Συστήματα Υπολογισμού Υψηλών Επιδόσεων

Καλημέρης Μάρκος 1659

Lab 3

Ερώτημα 1

Ο kernel για τις στήλες και ο kernel για τις γραμμές χρησιμοποιεί τετραγωνικό tile επομένως στην shared memory έρχεται ένα σημείο το οποίο το φέρνει κάθε thread από τη global memory. Για τα υπόλοιπα σημεία που δεν υπάρχουν στη shared χρησιμοποιούμε την global memory απευθείας.

Ερώτημα 2

Γραμμές:

Registers: 23

Constant Memory: 1024 bytes

Shared Memory: 49152 bytes

Στήλες:

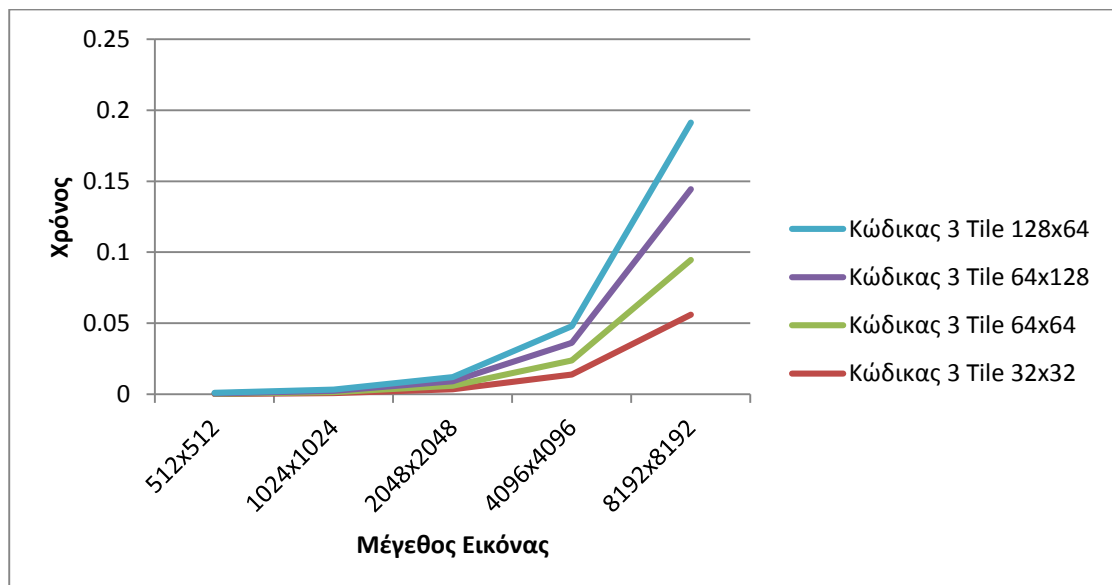
Registers: 28

Constant Memory: 1024 bytes

Shared Memory: 40960 bytes

Ερώτημα 3

Οι kernel για τις γραμμές και για τις στήλες χρησιμοποιούν Block 32x32, ενώ το tile είναι πολλαπλάσιο του μεγέθους του Block και καθορίζεται κατά τη μεταγλώττιση. Κάθε thread φέρνει $TILE_HEIGHT_Y * TILE_WIDTH_X$ σημεία, τα οποία εκφράζουν το πόσες φορές το tile είναι πολλαπλάσιο του block. Θεωρώντας φίλτρο 16x16 στον kernel των γραμμών το κάθε thread φέρνει και $TILE_HEIGHT_Y$ σημεία του padding και αντίστοιχα στις στήλες $TILE_WIDTH_X$ σημεία του padding. Φέρνουμε έτσι στη shared memory ό,τι χρειαζόμαστε νωρίτερα. Στην constant memory αποθηκεύεται το φίλτρο και η μνήμη cache έχει οριστεί στα 48KB. (Σε μέγεθος tile 128x128 δεν επαρκούσαν οι πόροι της μνήμης για αυτό και δεν χρησιμοποιήθηκε).



Ερώτημα 4

•

Kernel Γραμμών: tile $128 \times 64 + \text{padding} = 128 \times 96$ και θεωρούμε τη μέγιστη ανάλυση εικόνας 8192×8192 για να έχουμε και τον μέγιστο αριθμό των block.

Κάθε thread εκτελεί $2 * (1+4) + 2 * 4 * 33 = 274$ αναγνώσεις εικόνας και η δεύτερη πρόσθεση είναι οι αναγνώσεις του φίλτρου από την constant δηλαδή $2 * 4 * 33 = 264$ αναγνώσεις.

Για τη μνήμη global έχουμε $2 * (1+4) = 10$, για τη shared και την constant $2 * 4 * 33 = 264$.

Ο συνολικός αριθμός των threads είναι $32 \times 32 \times 8192 = 8.388.608$

Και ο συνολικός αριθμός προσπελάσεων είναι $8.388.608 * 274 = 2.298.478.592$

Για κάθε στοιχείο της εικόνας έχουμε:

Σύνολο: $2.298.478.592 / (8192 * 8192) = 34,25$

Global: 1,25

Shared: 33

Shared/Global = 26,4

Για κάθε στοιχείο του φίλτρου:

Σύνολο: Συνολικές Προσπελάσεις / $(16 * 16) = 8650752$

Kernel Στηλών: tile $128 \times 64 + \text{padding} = 160 \times 64$ και θεωρούμε τη μέγιστη ανάλυση εικόνας 8192×8192 για να έχουμε και τον μέγιστο αριθμό των block.

Κάθε thread εκτελεί $4 * (1+2) + 2 * 4 * 33 = 276$ αναγνώσεις εικόνας και η δεύτερη πρόσθεση είναι οι αναγνώσεις του φίλτρου από την constant δηλαδή $2 * 4 * 33 = 264$ αναγνώσεις.

Για τη μνήμη global έχουμε $4 * (1+2) = 12$, για τη shared και την constant $2 * 4 * 33 = 264$.

Ο συνολικός αριθμός των threads είναι $32 \times 32 \times 8192 = 8.388.608$

Και ο συνολικός αριθμός προσπελάσεων είναι $8.388.608 * 274 = 2.298.478.592$

Για κάθε στοιχείο της εικόνας έχουμε:

Σύνολο: $2.298.478.592 / (8192 * 8192) = 34,5$

Global: 1,5

Shared: 33

Shared/Global = 22

Για κάθε στοιχείο του φίλτρου:

Σύνολο: Συνολικές Προσπελάσεις / $(16 * 16) = 8650752$

•

Kernel Γραμμών:

Προσπελάσεις στη global memory: $2 * (1+4) = 10$

Πράξεις κινητής υποδιαστολής: $2 * 4 * 33 * 2 = 528$

Λόγος: 52,8

Kernel Στηλών:

Προσπελάσεις στη global memory: $4 \cdot (1+2) = 12$

Πράξεις κινητής υποδιαστολής: $2 \cdot 4 \cdot 33 \cdot 2 = 528$

Λόγος: 44

- Επειδή χρησιμοποιούμε τη shared memory σαν cache έχουμε αρκετά βελτιωμένο χρόνο εκτέλεσης αφού τα στοιχεία της εικόνας δε χρειάζεται να τα διαβάσουμε πλέον από την global memory. Χρησιμοποιώντας την constant memory για να αποθηκεύσουμε το φίλτρο καταφέρνουμε να μειώσουμε κι άλλο το χρόνο εκτέλεσης γιατί τα στοιχεία διαβάζονται από την constant και όχι από τη global.

Ερώτημα 5

Λόγω περιορισμού της shared memory παρόλο που έχουμε μέγεθος block 32x32 το μέγεθος του tile δε μπορεί να υπερβεί τα 96x64 στον kernel των γραμμών και το 64x96 στον kernel των στηλών.

