Project Presentation

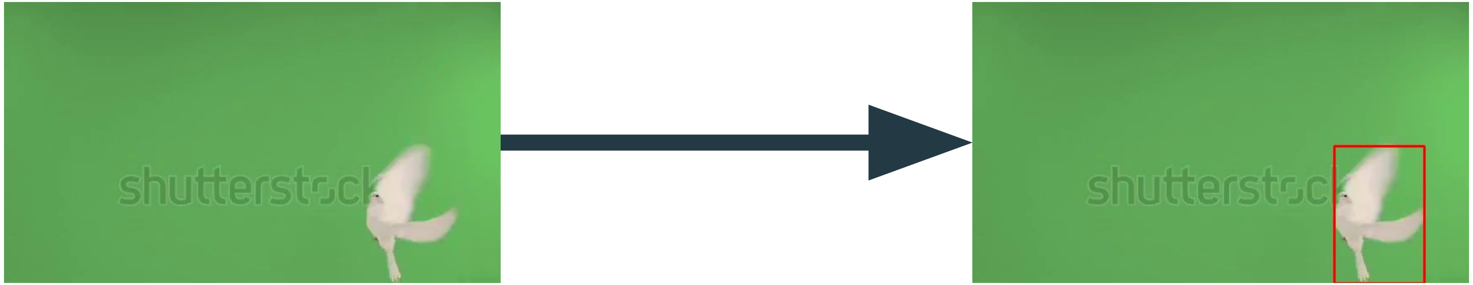
Moving Object Detector

CS677 Parallel Computing for Many Core Processors

Adrien HUET

Problem Description

- Implement a program capable of delimiting an moving object in a video input



Algorithm

(1) Reserve the first frame $g0$ of the video as a background reference for further frames.

(2) Convert both the background $g0$ and the frame to process $g1$ into grayscale.

(3) Smooth the two images $g0$ and $g1$ (e.g. with a Gaussian filter).

(4) Compute the difference $d = |g0 - g1|$ between the two images.

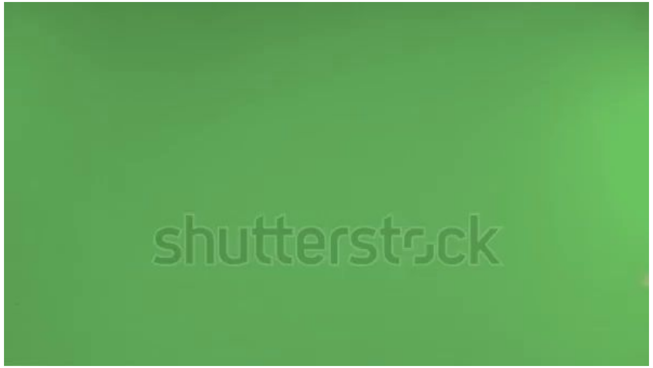
(5) Perform morphological closing/opening to remove non-meaningful artifacts.

(6) Threshold the image d and keep only the connected components (label propagation).

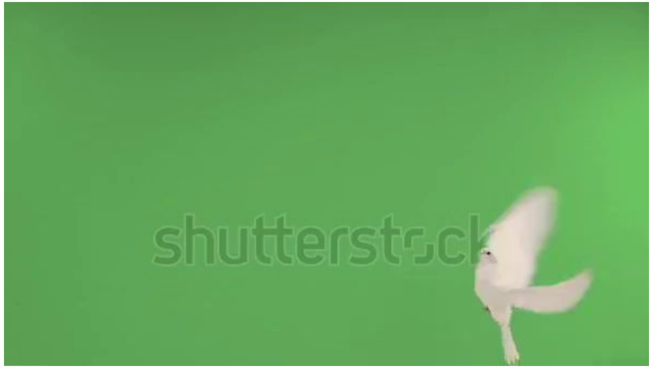
(7) Compute and output the bounding boxes.

(8) (Render new video with boxes drawn onto each frame)

Algorithm

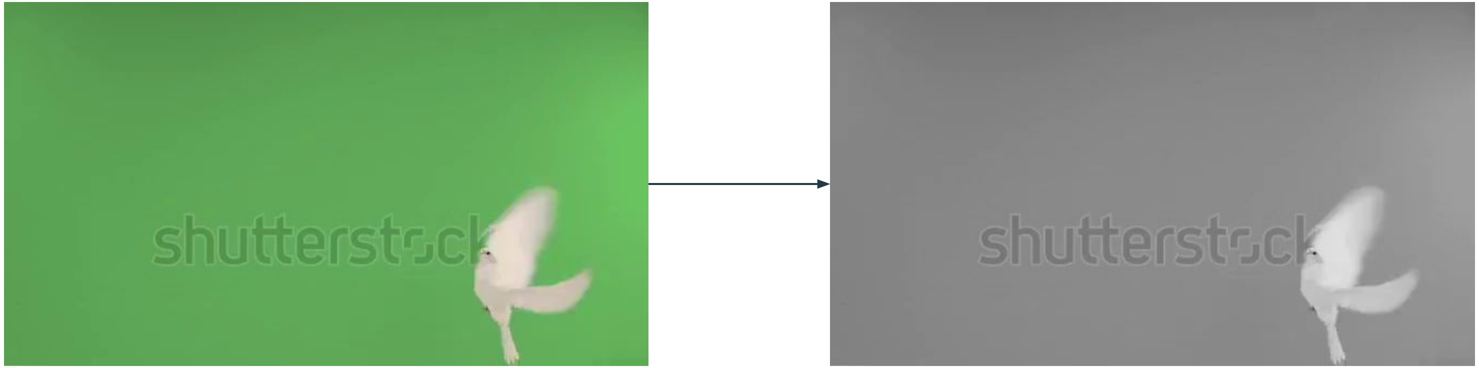


Background



Frame

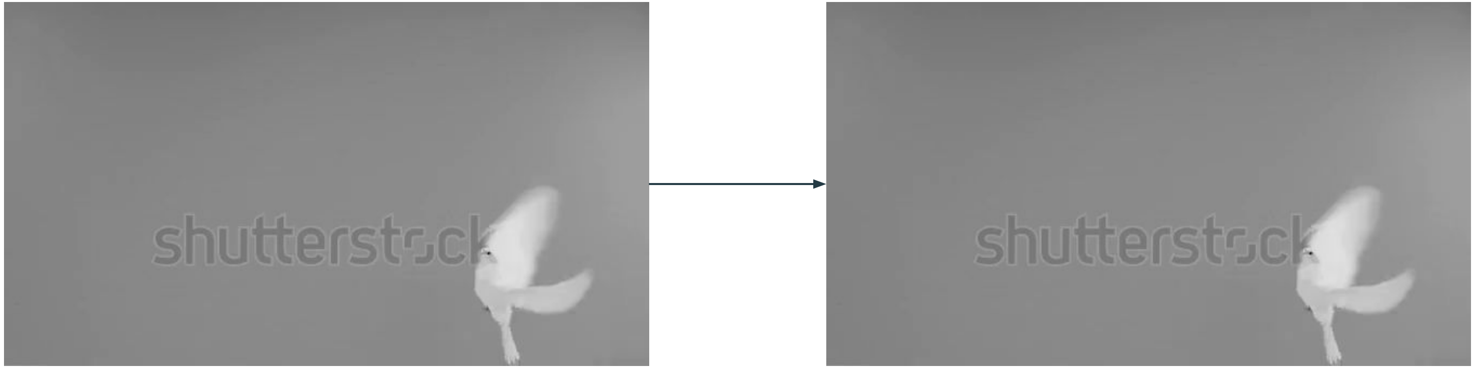
Algorithm



Frame

Grayscale

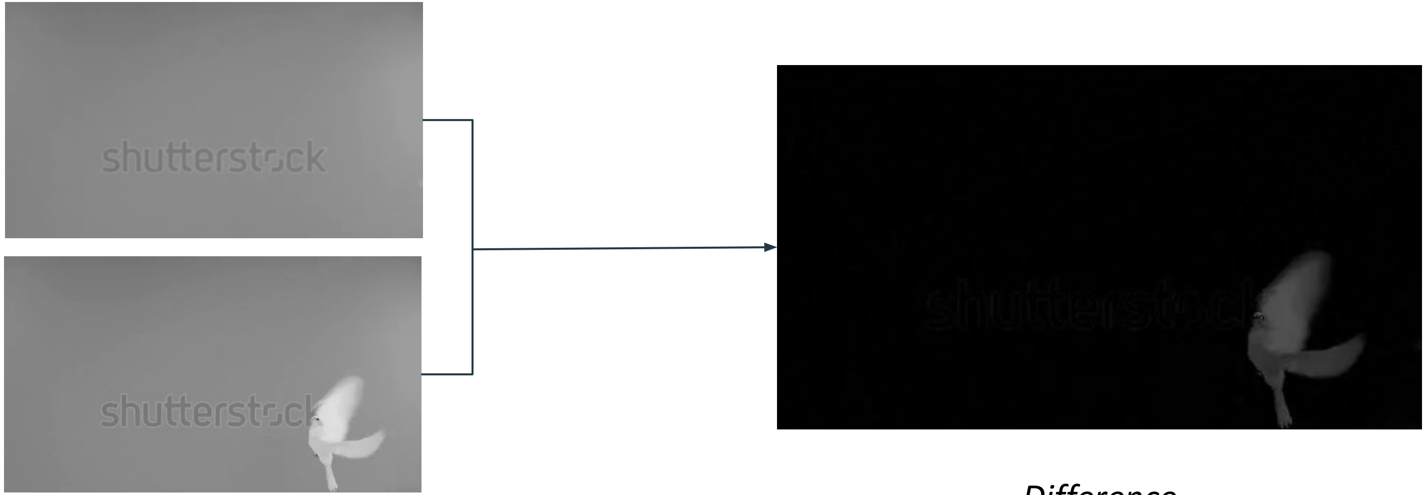
Algorithm



Grayscale

Blur

Algorithm



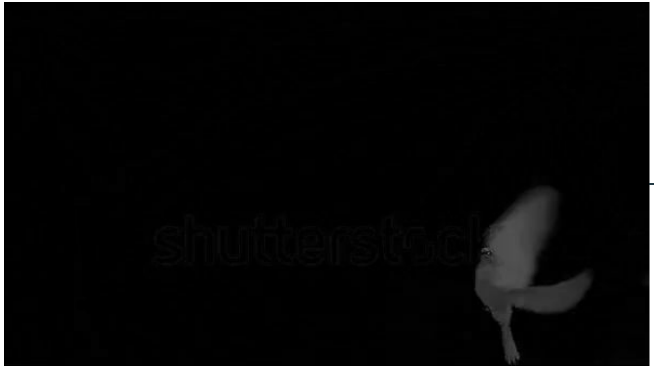
Difference

Algorithm

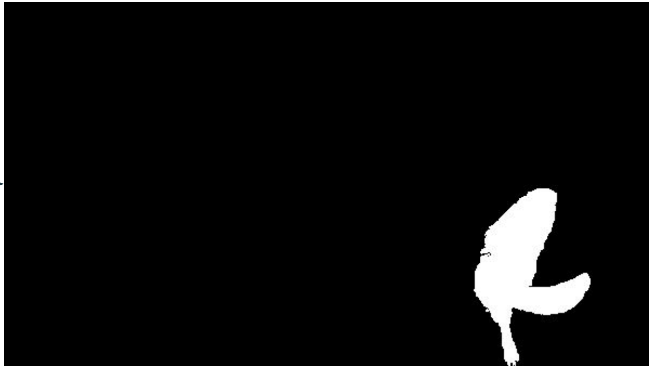
Difference

Morphological opening

Algorithm

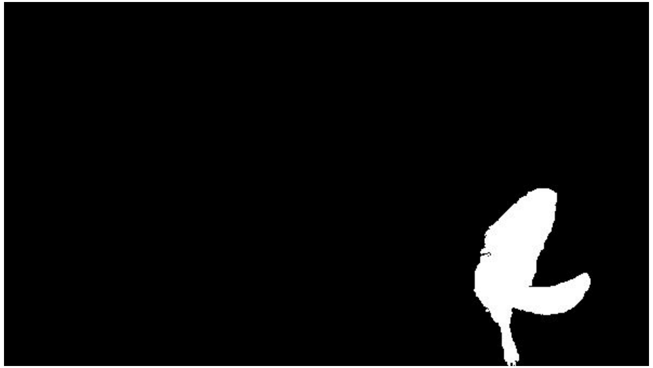


Morphological opening

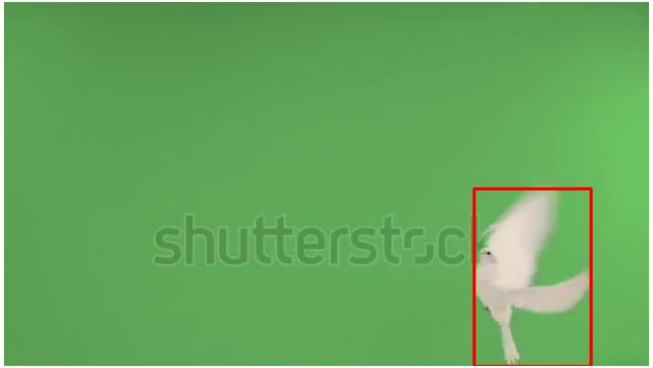


Threshold

Algorithm



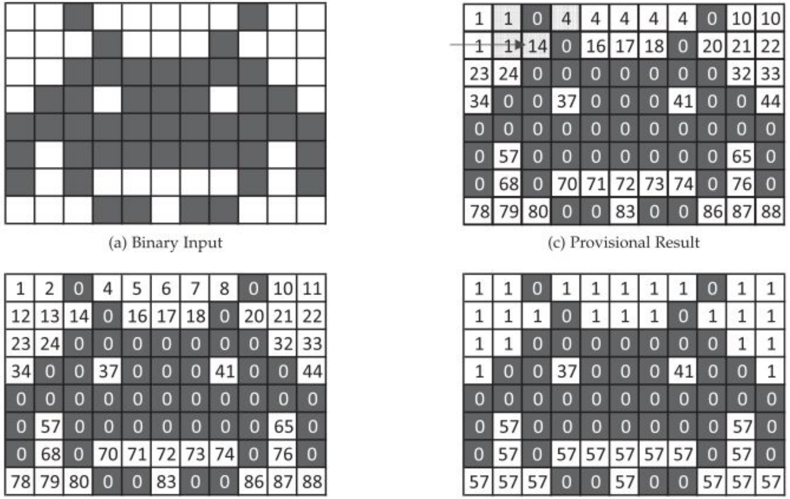
Threshold



Symbolic CC
Labelled Image
+
Bounding Box

Connected Components Labelling - Union Find

"Optimizing GPU-Based Connected Components Labeling Algorithms", Stefano Allegretti, Federico Bolelli *et al.*, 2018, IPAS
"Optimized Block-Based Algorithms to Label Connected Components on GPUs", Stefano Allegretti, Federico Bolelli *et al.,* 2020, IEEE



□	□	■	□	□	□	□	□	■	□	□
□	□	□	■	□	□	□	■	□	□	□
□	□	■	■	■	■	■	■	■	□	□
□	■	■	□	■	■	■	□	■	■	□
■	■	■	■	■	■	■	■	■	■	■
■	□	■	■	■	■	■	■	■	□	■
■	□	■	□	□	□	□	□	■	□	■
□	□	□	■	■	□	■	■	□	□	□

(a) Binary Input

1	1	0	4	4	4	4	4	0	10	10
1	1	14	0	16	17	18	0	20	21	22
23	24	0	0	0	0	0	0	0	32	33
34	0	0	37	0	0	0	41	0	0	44
0	0	0	0	0	0	0	0	0	0	0
0	57	0	0	0	0	0	0	0	65	0
0	68	0	70	71	72	73	74	0	76	0
78	79	80	0	0	83	0	0	86	87	88

(c) Provisional Result

1	2	0	4	5	6	7	8	0	10	11
12	13	14	0	16	17	18	0	20	21	22
23	24	0	0	0	0	0	0	0	32	33
34	0	0	37	0	0	0	41	0	0	44
0	0	0	0	0	0	0	0	0	0	0
0	57	0	0	0	0	0	0	0	65	0
0	68	0	70	71	72	73	74	0	76	0
78	79	80	0	0	83	0	0	86	87	88

(b) Output Initialization

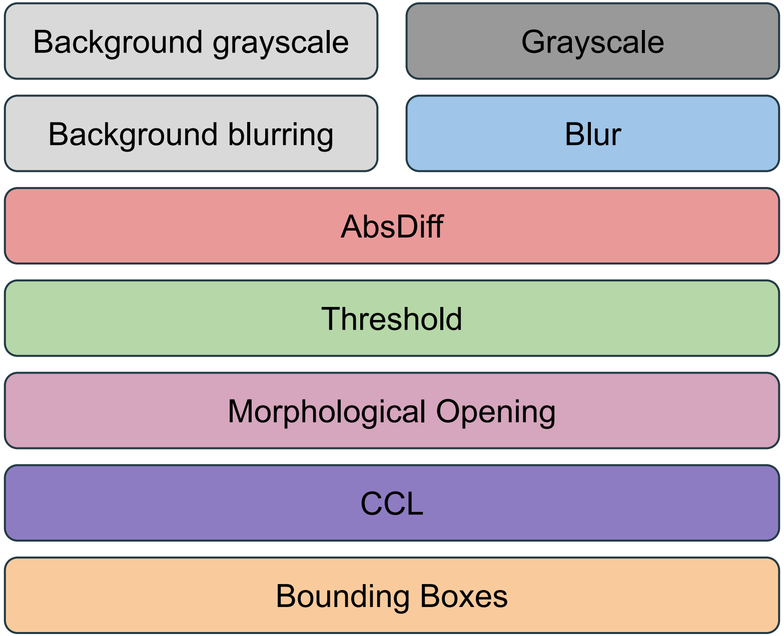
1	1	0	1	1	1	1	1	0	1	1
1	1	1	0	1	1	1	0	1	1	1
1	1	0	0	0	0	0	0	0	1	1
1	0	0	37	0	0	0	41	0	0	1
0	0	0	0	0	0	0	0	0	0	0
0	57	0	0	0	0	0	0	0	57	0
0	57	0	57	57	57	57	57	0	57	0
57	57	57	0	0	57	0	0	57	57	57

(d) Output Labels

Suitability for GPU acceleration

- All steps can be executed on the GPU
- Morphological operations, blurring and finding connected components are the most expensive
- Low data footprint: predicted 1 load for background and 1 for each frame

GPU Implementation



Background grayscale

Grayscale

Background blurring

Blur

AbsDiff

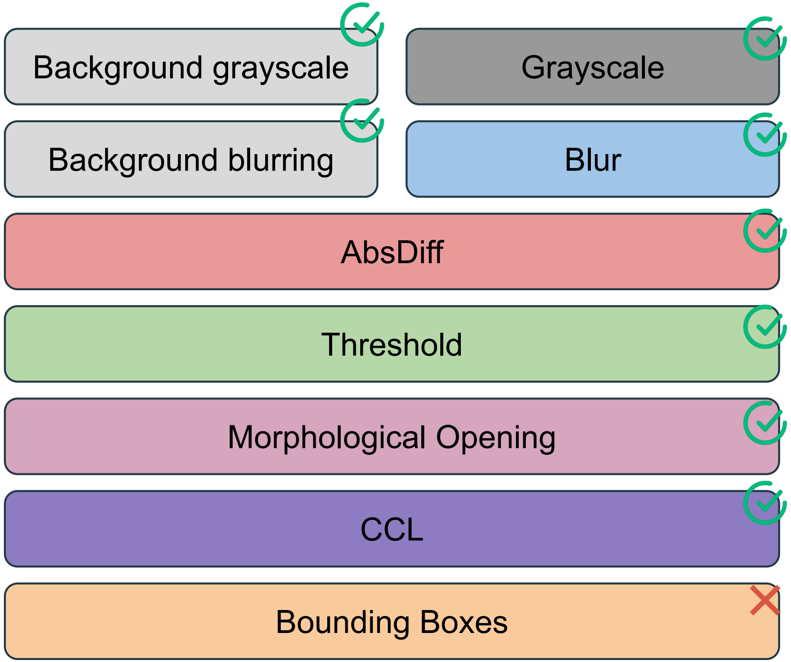
Threshold

Morphological Opening

CCL

Bounding Boxes

GPU Implementation



Background grayscale ✓

Grayscale ✓

Background blurring ✓

Blur ✓

AbsDiff ✓

Threshold ✓

Morphological Opening ✓

CCL ✓

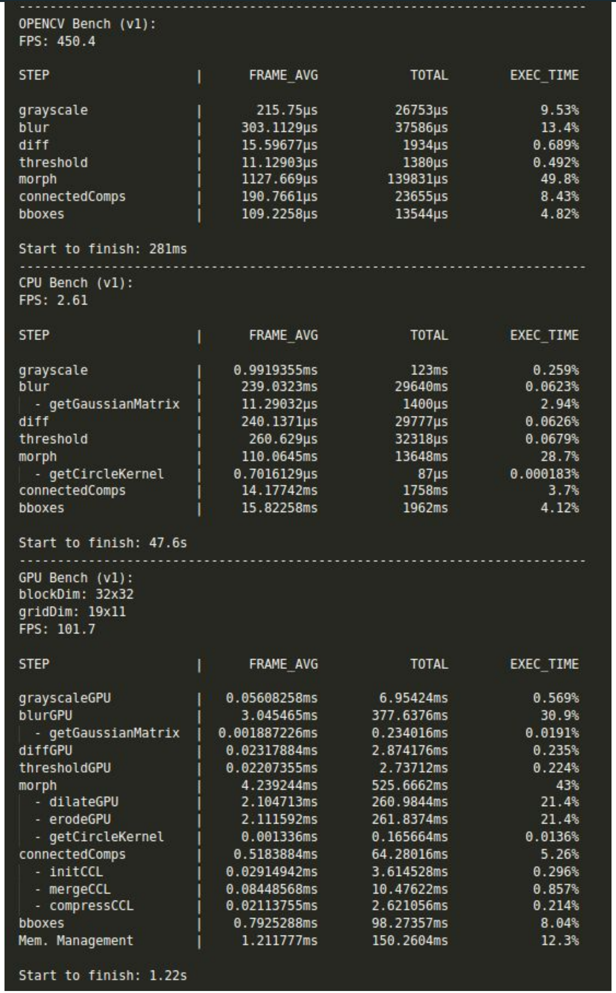
Bounding Boxes ✗

GPU V1.0

Very naive implementation

Allocation and deallocation for each frame

Background and kernels are processed for each frame

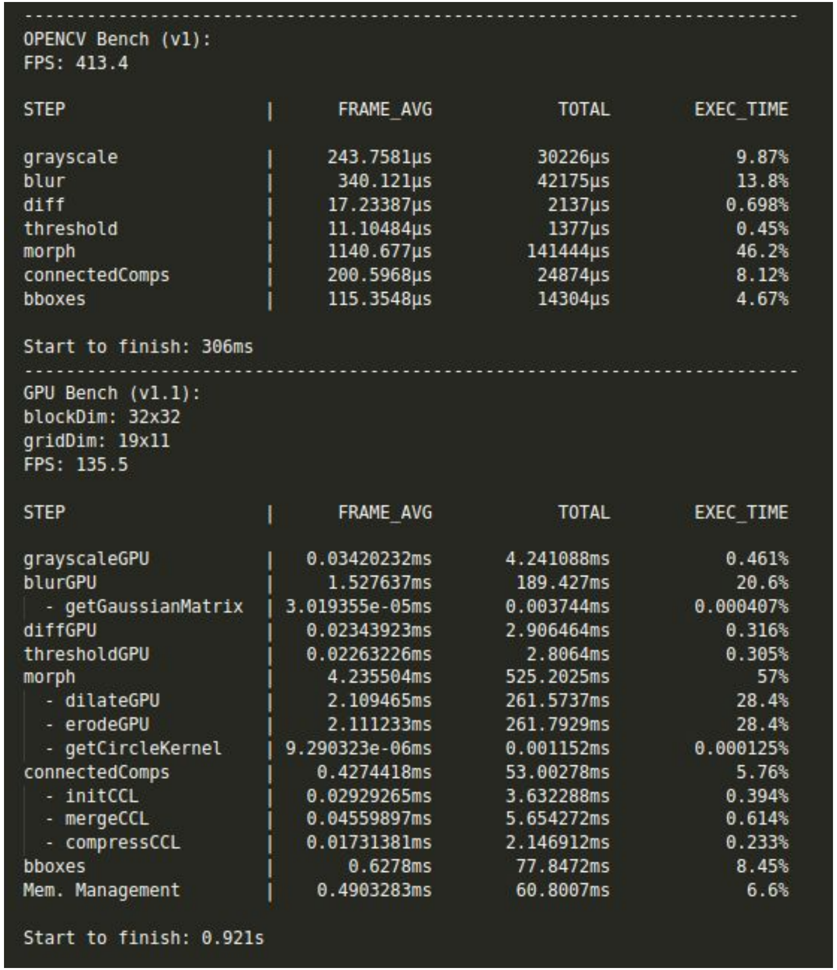


```
--------------------------------------------------------------------------
OPENCV Bench (v1):
FPS: 450.4

STEP                   |       FRAME_AVG            TOTAL        EXEC_TIME

grayscale              |        215.75µs          26753µs            9.53%
blur                   |      303.1129µs          37586µs            13.4%
diff                   |      15.59677µs           1934µs           0.689%
threshold              |      11.12903µs           1380µs           0.492%
morph                  |      1127.669µs         139831µs            49.8%
connectedComps         |      190.7661µs          23655µs            8.43%
bboxes                 |      109.2258µs          13544µs            4.82%

Start to finish: 281ms
--------------------------------------------------------------------------
CPU Bench (v1):
FPS: 2.61

STEP                   |       FRAME_AVG            TOTAL        EXEC_TIME

grayscale              |     0.9919355ms            123ms           0.259%
blur                   |      239.0323ms          29640ms          0.0623%
  - getGaussianMatrix  |      11.29032µs           1400µs            2.94%
diff                   |      240.1371µs          29777µs          0.0626%
threshold              |       260.629µs          32318µs          0.0679%
morph                  |      110.0645ms          13648ms            28.7%
  - getCircleKernel    |     0.7016129µs             87µs        0.000183%
connectedComps         |      14.17742ms           1758ms             3.7%
bboxes                 |      15.82258ms           1962ms            4.12%

Start to finish: 47.6s
--------------------------------------------------------------------------
GPU Bench (v1):
blockDim: 32x32
gridDim: 19x11
FPS: 101.7

STEP                   |       FRAME_AVG            TOTAL        EXEC_TIME

grayscaleGPU           |    0.05608258ms        6.95424ms           0.569%
blurGPU                |      3.045465ms       377.6376ms            30.9%
  - getGaussianMatrix  |   0.001887226ms       0.234016ms          0.0191%
diffGPU                |    0.02317884ms       2.874176ms           0.235%
thresholdGPU           |    0.02207355ms        2.73712ms           0.224%
morph                  |      4.239244ms       525.6662ms              43%
  - dilateGPU          |      2.104713ms       260.9844ms            21.4%
  - erodeGPU           |      2.111592ms       261.8374ms            21.4%
  - getCircleKernel    |      0.001336ms       0.165664ms          0.0136%
connectedComps         |     0.5183884ms       64.28016ms            5.26%
  - initCCL            |    0.02914942ms       3.614528ms           0.296%
  - mergeCCL           |    0.08448568ms        10.47622ms          0.857%
  - compressCCL        |    0.02113755ms       2.621056ms           0.214%
bboxes                 |     0.7925288ms       98.27357ms            8.04%
Mem. Management        |      1.211777ms       150.2604ms            12.3%

Start to finish: 1.22s
```

GPU V1.1

Changes:
- Background is processed only once
- Memory is allocated and free only once (based on frame dimensions)

Effect:
Only 2 Memcpy per frame and overhaul performance improvement
⅓ of memory management time
~+30 FPS



```
--------------------------------------------------------------------------
OPENCV Bench (v1):
FPS: 413.4

STEP                   |      FRAME_AVG            TOTAL        EXEC_TIME

grayscale              |     243.7581µs          30226µs            9.87%
blur                   |      340.121µs          42175µs            13.8%
diff                   |     17.23387µs           2137µs           0.698%
threshold              |     11.10484µs           1377µs            0.45%
morph                  |     1140.677µs         141444µs            46.2%
connectedComps         |     200.5968µs          24874µs            8.12%
bboxes                 |     115.3548µs          14304µs            4.67%

Start to finish: 306ms
--------------------------------------------------------------------------
GPU Bench (v1.1):
blockDim: 32x32
gridDim: 19x11
FPS: 135.5

STEP                   |      FRAME_AVG            TOTAL        EXEC_TIME

grayscaleGPU           |   0.03420232ms       4.241088ms           0.461%
blurGPU                |     1.527637ms        189.427ms            20.6%
  - getGaussianMatrix  | 3.019355e-05ms       0.003744ms        0.000407%
diffGPU                |   0.02343923ms       2.906464ms           0.316%
thresholdGPU           |   0.02263226ms         2.8064ms           0.305%
morph                  |     4.235504ms       525.2025ms              57%
  - dilateGPU          |     2.109465ms       261.5737ms            28.4%
  - erodeGPU           |     2.111233ms       261.7929ms            28.4%
  - getCircleKernel    | 9.290323e-06ms       0.001152ms        0.000125%
connectedComps         |    0.4274418ms       53.00278ms            5.76%
  - initCCL            |   0.02929265ms       3.632288ms           0.394%
  - mergeCCL           |   0.04559897ms       5.654272ms           0.614%
  - compressCCL        |   0.01731381ms       2.146912ms           0.233%
bboxes                 |       0.6278ms        77.8472ms            8.45%
Mem. Management        |    0.4903283ms        60.8007ms             6.6%

Start to finish: 0.921s
```

GPU V1.2

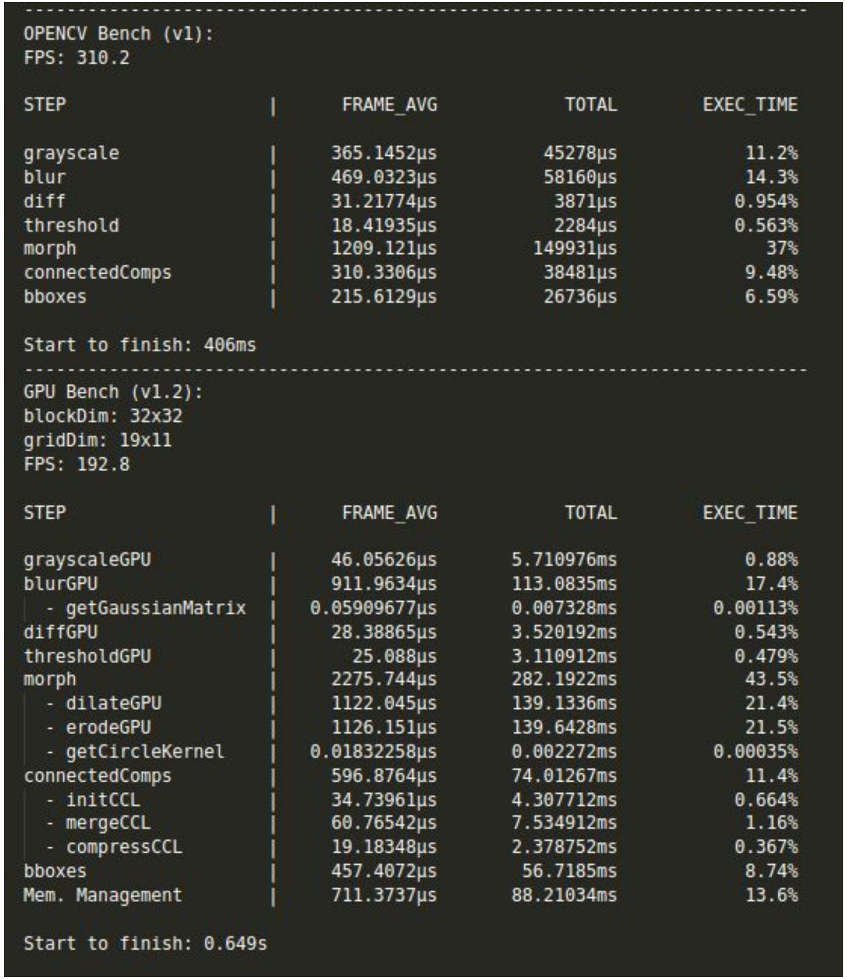
Changes:
- Shared memory tiling for blur and morph
- Input/Output tile version differs drastically
- Constant memory for kernel masks

Effect:

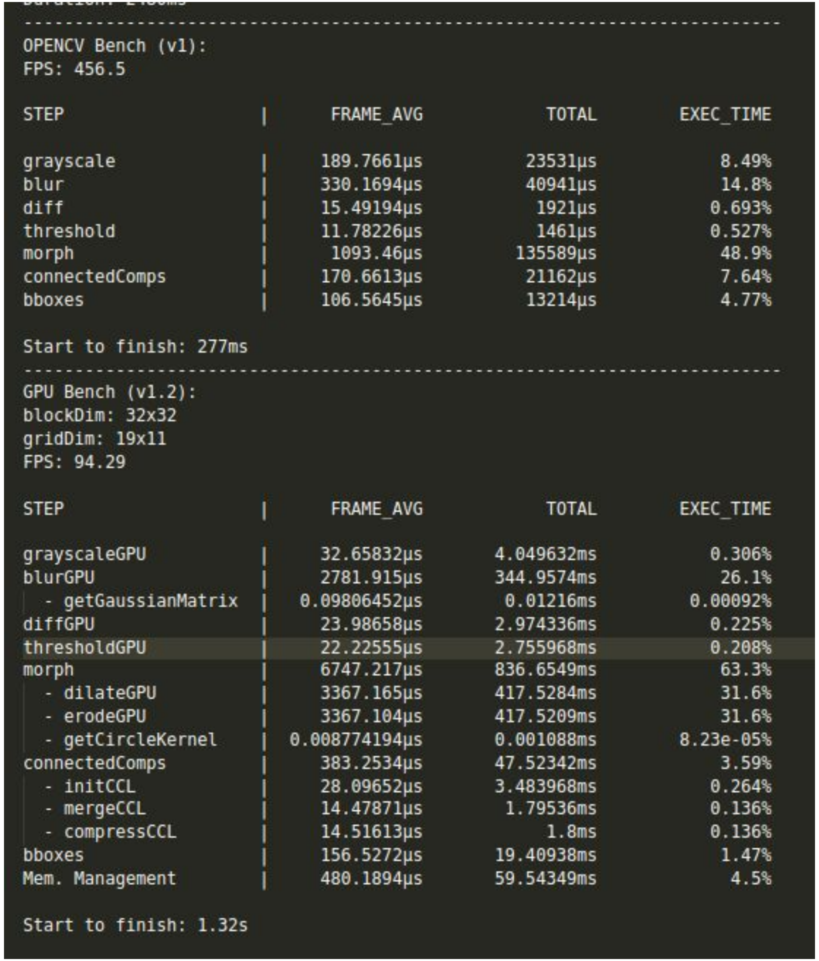
First sm version double exec time

Second version halved it

~+60 FPS



```
--------------------------------------------------------------------------
OPENCV Bench (v1):
FPS: 310.2

STEP                   |    FRAME_AVG          TOTAL      EXEC_TIME

grayscale              |   365.1452µs        45278µs          11.2%
blur                   |   469.0323µs        58160µs          14.3%
diff                   |   31.21774µs         3871µs         0.954%
threshold              |   18.41935µs         2284µs         0.563%
morph                  |   1209.121µs       149931µs            37%
connectedComps         |   310.3306µs        38481µs          9.48%
bboxes                 |   215.6129µs        26736µs          6.59%

Start to finish: 406ms
--------------------------------------------------------------------------
GPU Bench (v1.2):
blockDim: 32x32
gridDim: 19x11
FPS: 192.8

STEP                   |    FRAME_AVG          TOTAL      EXEC_TIME

grayscaleGPU           |   46.05626µs    5.710976ms          0.88%
blurGPU                |   911.9634µs    113.0835ms          17.4%
  - getGaussianMatrix  | 0.05909677µs    0.007328ms       0.00113%
diffGPU                |   28.38865µs    3.520192ms         0.543%
thresholdGPU           |     25.088µs    3.110912ms         0.479%
morph                  |   2275.744µs    282.1922ms          43.5%
  - dilateGPU          |   1122.045µs    139.1336ms          21.4%
  - erodeGPU           |   1126.151µs    139.6428ms          21.5%
  - getCircleKernel    | 0.01832258µs    0.002272ms       0.00035%
connectedComps         |   596.8764µs    74.01267ms          11.4%
  - initCCL            |   34.73961µs    4.307712ms         0.664%
  - mergeCCL           |   60.76542µs    7.534912ms          1.16%
  - compressCCL        |   19.18348µs    2.378752ms         0.367%
bboxes                 |   457.4072µs     56.7185ms          8.74%
Mem. Management        |   711.3737µs    88.21034ms          13.6%

Start to finish: 0.649s
```
]



```
----------------------------------------------------------------------------
OPENCV Bench (v1):
FPS: 456.5

STEP                   |       FRAME_AVG          TOTAL        EXEC_TIME

grayscale              |      189.7661µs        23531µs            8.49%
blur                   |      330.1694µs        40941µs            14.8%
diff                   |      15.49194µs         1921µs           0.693%
threshold              |      11.78226µs         1461µs           0.527%
morph                  |       1093.46µs       135589µs            48.9%
connectedComps         |      170.6613µs        21162µs            7.64%
bboxes                 |      106.5645µs        13214µs            4.77%

Start to finish: 277ms
----------------------------------------------------------------------------
GPU Bench (v1.2):
blockDim: 32x32
gridDim: 19x11
FPS: 94.29

STEP                   |       FRAME_AVG          TOTAL        EXEC_TIME

grayscaleGPU           |      32.65832µs     4.049632ms           0.306%
blurGPU                |      2781.915µs     344.9574ms            26.1%
 | - getGaussianMatrix |    0.09806452µs      0.01216ms          0.00092%
diffGPU                |      23.98658µs     2.974336ms           0.225%
thresholdGPU           |      22.22555µs     2.755968ms           0.208%
morph                  |      6747.217µs     836.6549ms            63.3%
 | - dilateGPU         |      3367.165µs     417.5284ms            31.6%
 | - erodeGPU          |      3367.104µs     417.5209ms            31.6%
 | - getCircleKernel   |   0.008774194µs     0.001088ms         8.23e-05%
connectedComps         |      383.2534µs     47.52342ms            3.59%
 | - initCCL           |      28.09652µs     3.483968ms           0.264%
 | - mergeCCL          |      14.47871µs      1.79536ms           0.136%
 | - compressCCL       |      14.51613µs          1.8ms           0.136%
bboxes                 |      156.5272µs     19.40938ms            1.47%
Mem. Management        |      480.1894µs     59.54349ms             4.5%

Start to finish: 1.32s
```

First version of the shared memory load

GPU V1.3

<u>Changes:</u>

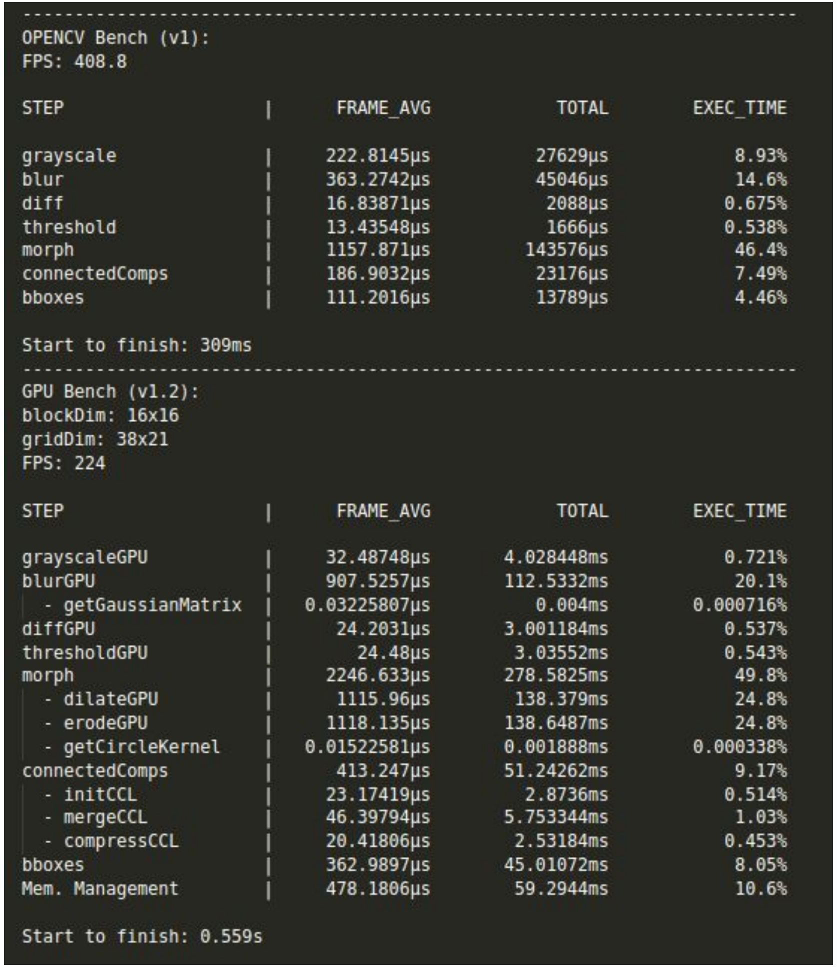
- Simple search for optimal block dimensions
- From 4 to 32-wide-blocks, 16 are the most efficient

<u>Effect:</u>

First sm version double exec time

Second version halved it

~+30 FPS



```
--------------------------------------------------------------------------
OPENCV Bench (v1):
FPS: 408.8

STEP                   |      FRAME_AVG            TOTAL        EXEC_TIME

grayscale              |    222.8145µs          27629µs            8.93%
blur                   |    363.2742µs          45046µs            14.6%
diff                   |    16.83871µs           2088µs           0.675%
threshold              |    13.43548µs           1666µs           0.538%
morph                  |    1157.871µs         143576µs            46.4%
connectedComps         |    186.9032µs          23176µs            7.49%
bboxes                 |    111.2016µs          13789µs            4.46%

Start to finish: 309ms
--------------------------------------------------------------------------
GPU Bench (v1.2):
blockDim: 16x16
gridDim: 38x21
FPS: 224

STEP                   |      FRAME_AVG            TOTAL        EXEC_TIME

grayscaleGPU           |    32.48748µs       4.028448ms           0.721%
blurGPU                |    907.5257µs       112.5332ms            20.1%
  - getGaussianMatrix  |  0.03225807µs          0.004ms        0.000716%
diffGPU                |     24.2031µs       3.001184ms           0.537%
thresholdGPU           |       24.48µs        3.03552ms           0.543%
morph                  |    2246.633µs       278.5825ms            49.8%
  - dilateGPU          |     1115.96µs        138.379ms            24.8%
  - erodeGPU           |    1118.135µs       138.6487ms            24.8%
  - getCircleKernel    |  0.01522581µs       0.001888ms        0.000338%
connectedComps         |     413.247µs       51.24262ms            9.17%
  - initCCL            |    23.17419µs         2.8736ms           0.514%
  - mergeCCL           |    46.39794µs       5.753344ms            1.03%
  - compressCCL        |    20.41806µs        2.53184ms           0.453%
bboxes                 |    362.9897µs       45.01072ms            8.05%
Mem. Management        |    478.1806µs        59.2944ms            10.6%

Start to finish: 0.559s
```

Memory Usage

All memory allocation at launch:

- 2x 3-channel image [width * height * sizeof(uchar3)] ⇒ original background and input, one is free before processing
- 3x single-channel frame [width * height * sizeof(uchar3)] ⇒ current bgd, frame and swap frame

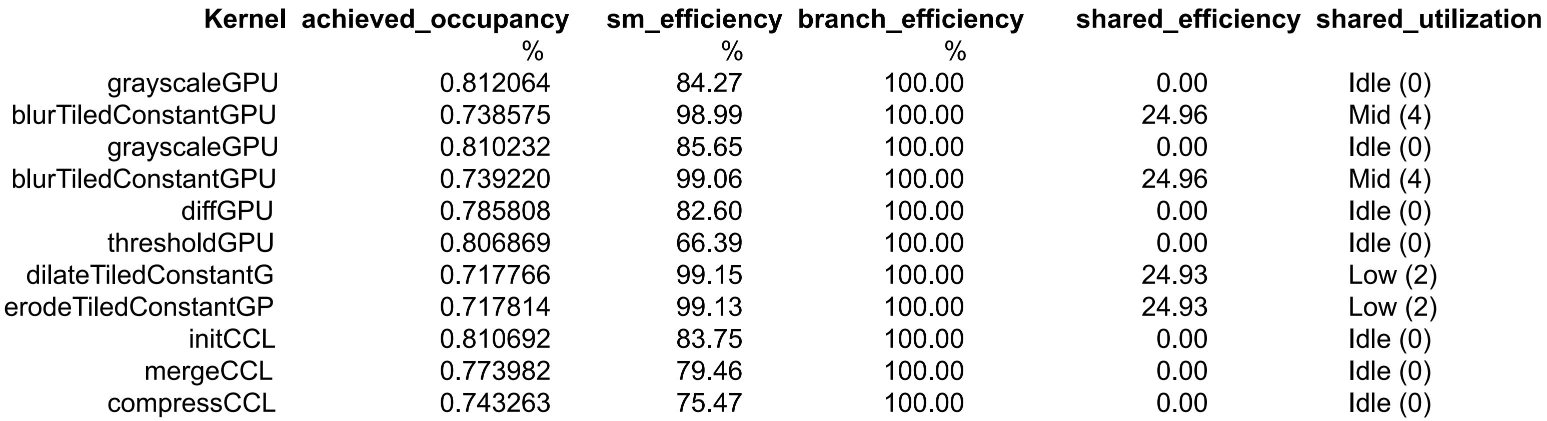
Memory bandwidth at launch

- 2x Memcpy to constant [ksize * ksize * sizeof(float|uchar)] ⇒ blur and morph kernels
- 1x Memcpy for background processing (3 frames == 1 image)

Memory bandwidth for each frame

- 1x Memcpy for original frame (~3 frames)
- 1x Memcpy to receive labelled frame [width * height * sizeof(**int**)]

nvprof - 16x16 blocks



Kernel	achieved_occupancy	sm_efficiency	branch_efficiency	shared_efficiency	shared_utilization
	%	%	%		
grayscaleGPU	0.812064	84.27	100.00	0.00	Idle (0)
blurTiledConstantGPU	0.738575	98.99	100.00	24.96	Mid (4)
grayscaleGPU	0.810232	85.65	100.00	0.00	Idle (0)
blurTiledConstantGPU	0.739220	99.06	100.00	24.96	Mid (4)
diffGPU	0.785808	82.60	100.00	0.00	Idle (0)
thresholdGPU	0.806869	66.39	100.00	0.00	Idle (0)
dilateTiledConstantG	0.717766	99.15	100.00	24.93	Low (2)
erodeTiledConstantGP	0.717814	99.13	100.00	24.93	Low (2)
initCCL	0.810692	83.75	100.00	0.00	Idle (0)
mergeCCL	0.773982	79.46	100.00	0.00	Idle (0)
compressCCL	0.743263	75.47	100.00	0.00	Idle (0)

What is well-optimized

- **Memory coalescing**, accesses are always coalesced
- **Memory communication** unless bboxes generation can be done on the GPU
- **Kernel launch overhead**, all kernels are executed sequentially, and there is not much to be done here

Future versions

- Shared memory prefetching for the blur and morph kernels
- Loop unrolling, specifically for memory loading
- Bounding boxes on GPU

Thank you for your attention!