Implementation of an Algorithm for Fast Down-Scale Transcoding of Compressed Video on the Itanium ™

Sumit Roy and Bo Shen
Client and Media Systems Laboratory
Hewlett-Packard Laboratories
Palo Alto, CA 94304

Overview

- Motivation
- Transcoding Algorithm
- Experiments
- Conclusion

Motivation

- High Quality Video being produced ($720 \times 480, 2 - 8\ Mbps$).
- Thin Mobile Clients becoming popular (Small screen, $128 - 384\ kbps$).
- Provide rich media experience.
- Handle live streams.

Conventional Down Scaling

- Decode to Pixel domain.
- Subsample to desired resolution.
- Re-encode to standard format.
- Too expensive !.

Compressed Domain Transcoding

- New Motion vector estimated from existing.
- Output Macroblock coding type based on input MBs.
- Quantization parameter derived using DCT coefficient activity from input MBs.

Profiling Output -$O3$

% time	seconds	calls	name
26.78	25.37	1188000	fdct
13.91	13.18		__divdi3
11.21	10.62	5537394	recon_comp
8.31	7.87	17843632	idctcol
4.57	4.33	31550362	flushbits
4.54	4.30	991488	quant_non_intra
4.43	4.20	600	getMBs
4.20	3.98	2230454	addblock
2.97	2.81	17843632	idctrow
2.38	2.25	600	down2PixelPic

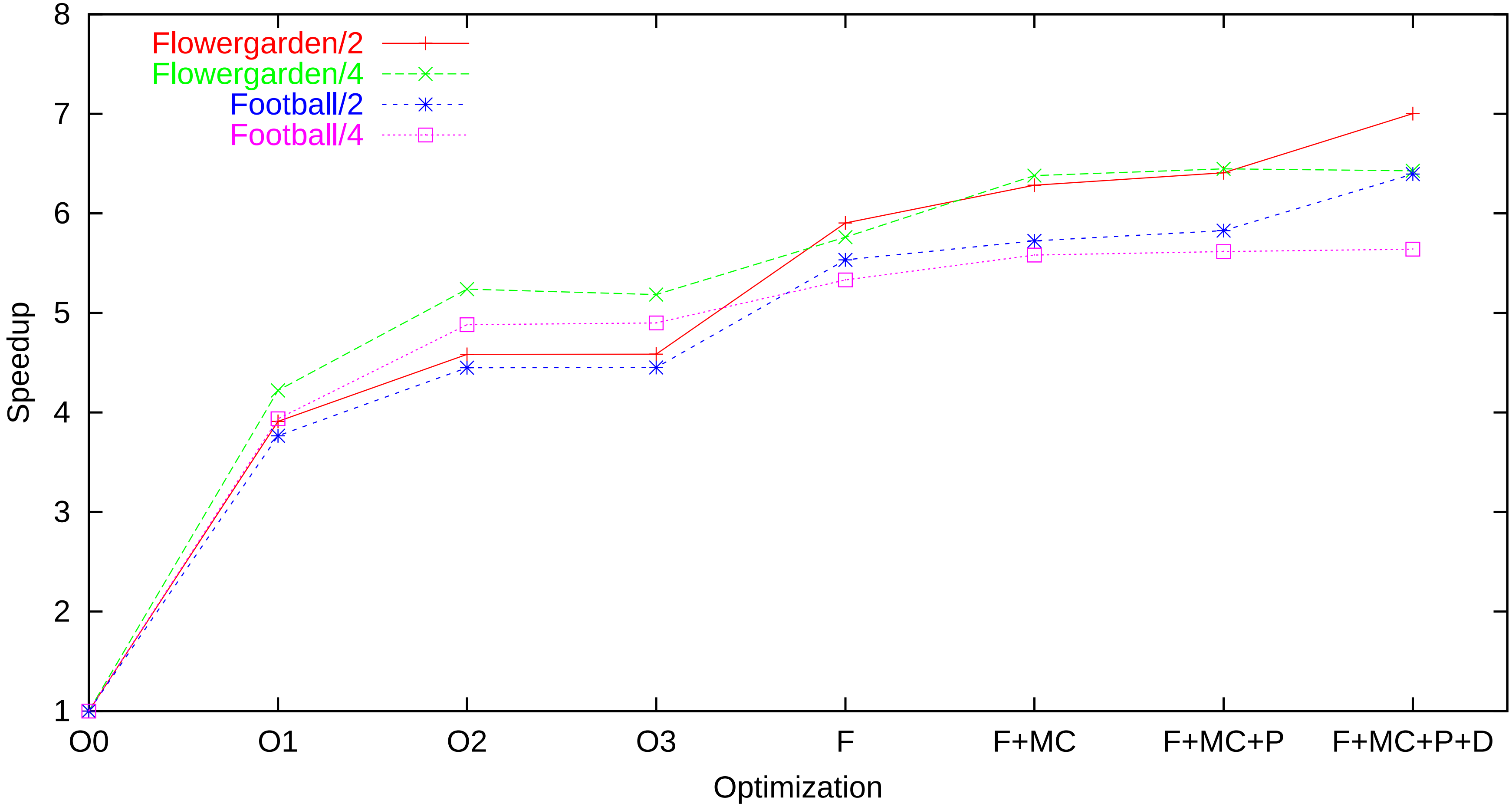
Multimedia Optimizations

- F: Forward DCT (`fdct`).
- MC: Motion Compensation (`recon_comp`).
- P: Prediction code (`predict`).
- D: Downsampling (`down2PixelPic`).

Profiling Output Multimedia, -*O3*

% time	seconds	calls	name
19.17	11.91		__divdi3
11.95	7.43	17843632	idctcol
6.93	4.31	31550362	flushbits
6.70	4.17	991488	quant_non_intra
6.59	4.09	600	getMBs
6.37	3.96	2230454	addblock
4.56	2.84	17843632	idctrow
3.96	2.46		forward_dct_row1
3.08	1.91	1761505	MC_put_y
2.44	1.51	792000	var_sblk

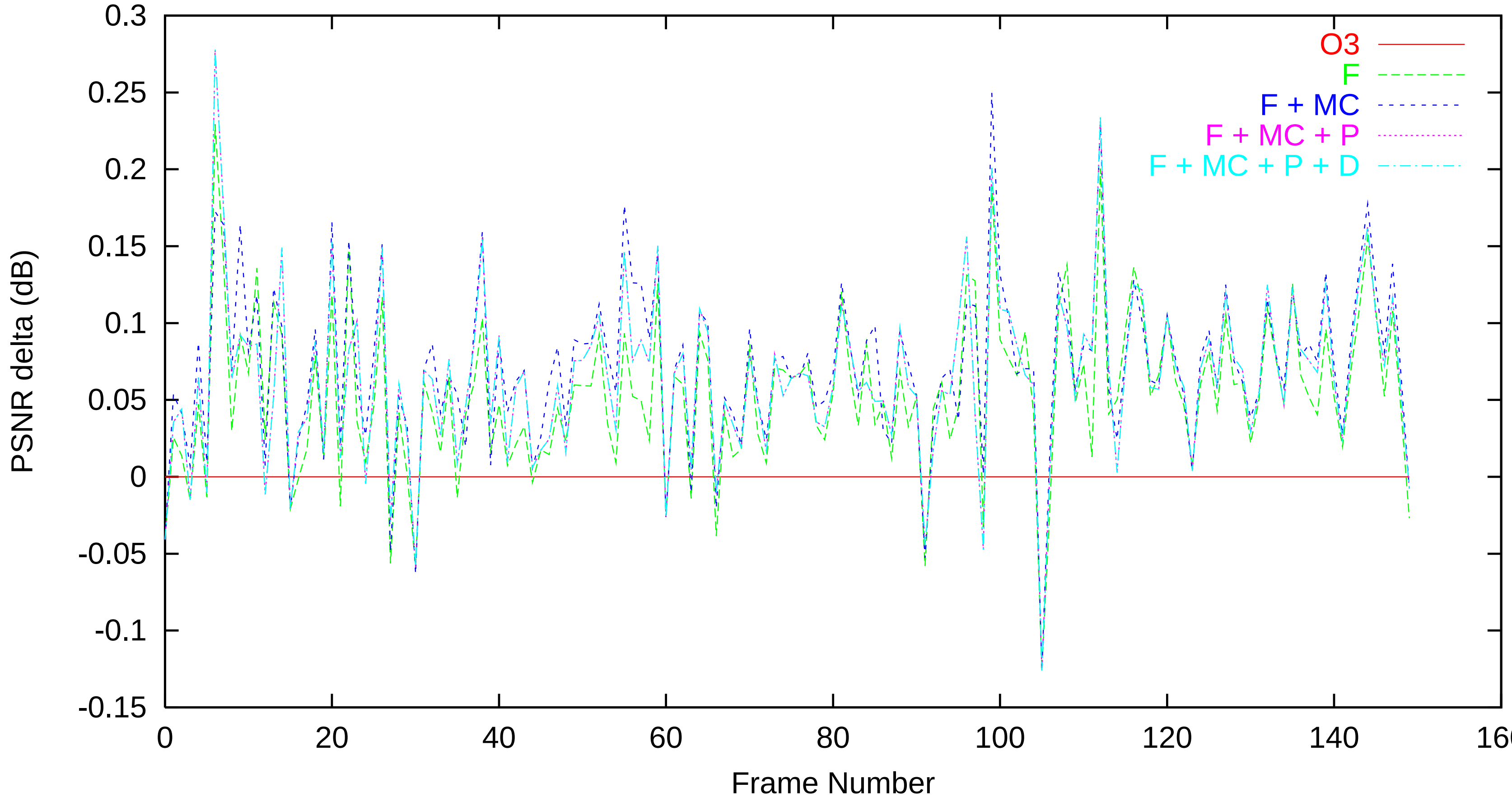
Speedup


Flowergarden/2
Flowergarden/4
Football/2
Football/4
Speedup
1
2
3
4
5
6
7
8
O0
O1
O2
O3
F
F+MC
F+MC+P
F+MC+P+D
Optimization

Percentage Improvement

Optimized Procedures	Baseline		
	O0	*O1*	*O3*
F	81.24 - 83.06	26.19 - 33.76	8.13 - 22.32
F+MC	82.08 - 84.32	29.49 - 37.75	12.24 - 27.01
F+MC+P	82.20 - 84.49	29.93 - 38.97	12.79 - 28.43
F+MC+P+D	82.27 - 85.72	30.23 - 44.15	13.15 - 34.51

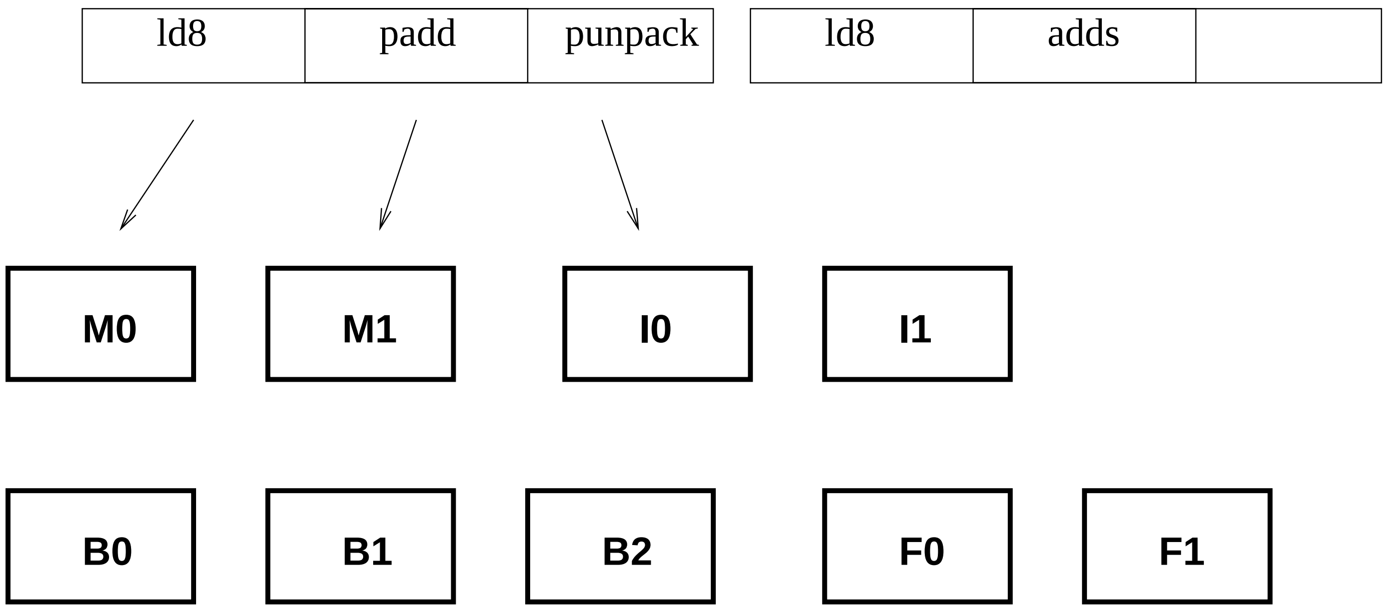
PSNR With Optimizations



Itanium Features

- Bundles of 3 instructions.
- Two bundle issue width.
- Multiple Functional Units.

Instruction Dispersal



FDCT Translated

1.	movd mm5, [INP+12];	1a.	adds r14 = 12, r32;;
		1b.	ld8 r25 = [r14]
2.	punpcklwd mm5, [INP+8];	2a.	adds r15 = 8, r32 ;;
		2b.	ld8 r29 = [r15];;
		2c.	unpack2.l r25 = r25, r29;;
3.	movq mm2, mm5;	3.	mov r22 = r25
4.	psrlq mm5, 32;	4.	shr.u r25 = r25, 32
5.	movq mm0, [INP];	5.	ld8 r30 = [r32];;
6.	punpcklwd mm5, mm2;	6.	unpack2.l r25 = r25, r22;;
7.	movq mm1, mm0;	7.	mov r21 = r30
8.	paddsw mm0, mm5;	8.	padd2.sss r30 = r30, r25;;
9.	psubsw mm1, mm5;	9.	psub2.sss r21 = r25, r21

FDCT Rescheduled

```
1a.   adds r14 = 12, r32;;
1b.   ld8 r25 = [r14]
2a.   adds r15 = 8, r32 ;;
2b.   ld8 r29 = [r15];;
2c.   unpack2.l r25 = r25, r29;;
3.    mov r22 = r25
4.    shr.u r25 = r25, 32
5.    ld8 r30 = [r32];;
6.    unpack2.l r25 = r25, r22;;
7.    mov r21 = r30
8.    padd2.sss r30 = r30, r25;;
9.    psub2.sss r21 = r25, r21
```

```
5.    ld8 r30 = [r32]
1a.   adds r14 = 12, r32
2a.   adds r15 = 8, r32;;
1b.   ld8 r25 = [r14]
2b.   ld8 r29 = [r15];;
2c.   unpack2.l r25 = r25, r29;;
4.    shr.u r22 = r25, 32;;
6.    unpack2.l r25 = r22, r25;;

9.    psub2.sss r21 = r25, r30
8.    padd2.sss r30 = r30, r25
```

Overall Results

Implementation			Performance
Original	Code Path	Latency	Metric
21684.6	19524.2	13746.9	CPU_CYCLES
17192.3	13832.2	12873.3	IA64_INST_RETIRED
11608.2	8620.1	5974.6	EXPL_STOPS_DISPERSED
13693.3	10329.2	7453.9	ALL_STOPS_DISPERSED
8258.3	5911.5	5058.4	DEPENDENCY_ALL_CYCLE
1469.9	2107.2	1611.2	DEPENDENCY_SCOREBOARD_CYCLE
3105.6	3193.0	3144.8	MEMORY_CYCLE
6348.7	3789.0	3549.1	NOPS_RETIRED
2660.9	2276.2	2338.0	UNSTALLED_BACKEND_CYCLE

Conclusion

- Algorithmic and Architectural optimizations required.
- Compiler has limitations (no multimedia code).
- Careful scheduling provides best results.