CUDA Total Variation Denoise

赵申剑
陈蕾宇
徐文康

2015.01.16

Basic Model

- Given an image $Y \in R^{m \times n}$, the discrete version of the Rudin-Osher-Fatemi (ROF) model is:

$$\min_X \frac{1}{2}\|X - Y\|_F^2 + \lambda \|X\|_{TV}$$

$$\|X\|_{TV} = \sum_{i=1}^{m} \sum_{j=1}^{n} \left\| D_{i,j} X \right\|_2$$

Solution

- Serial Optimization via Lagrange
- ADMM (Alternating Direction Method of Multipliers) based method: FAD(Fast ADMM for TV Models)

ICML2014

A Highly Scalable Parallel Algorithm for Isotropic Total Variation Models

ADMM Approach

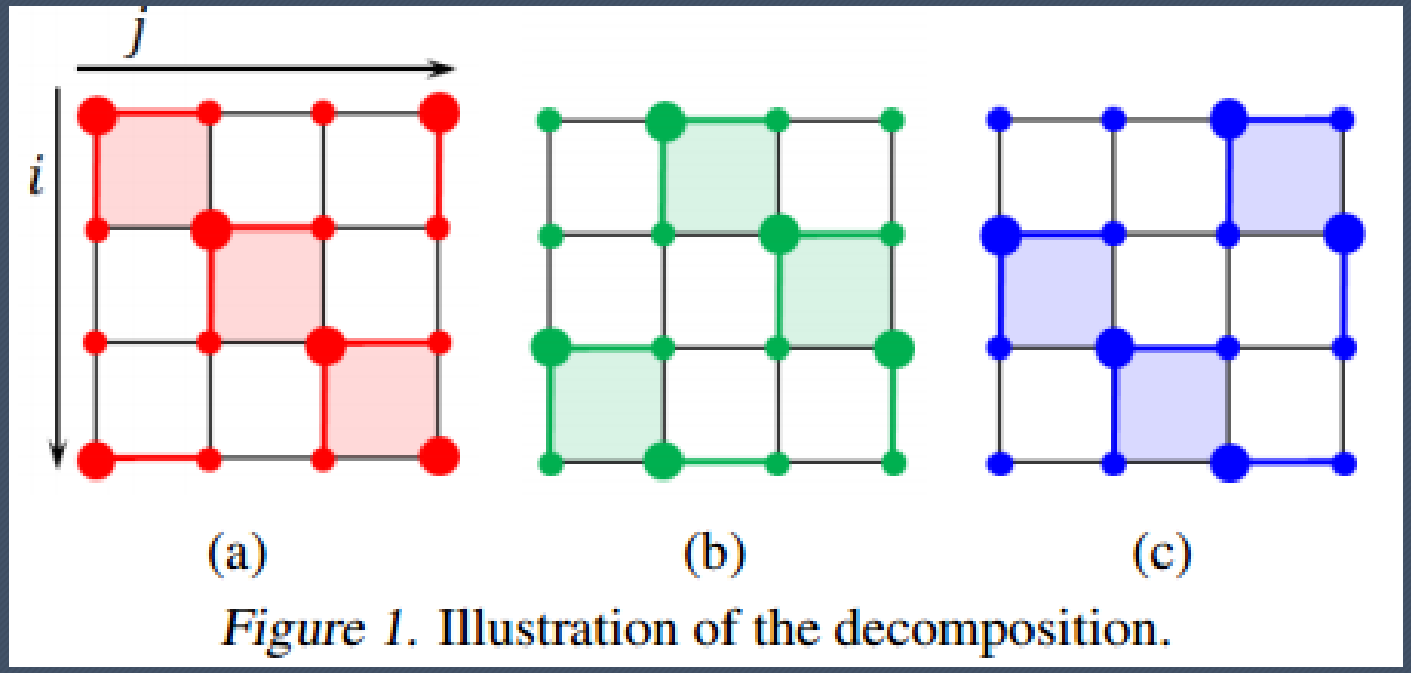
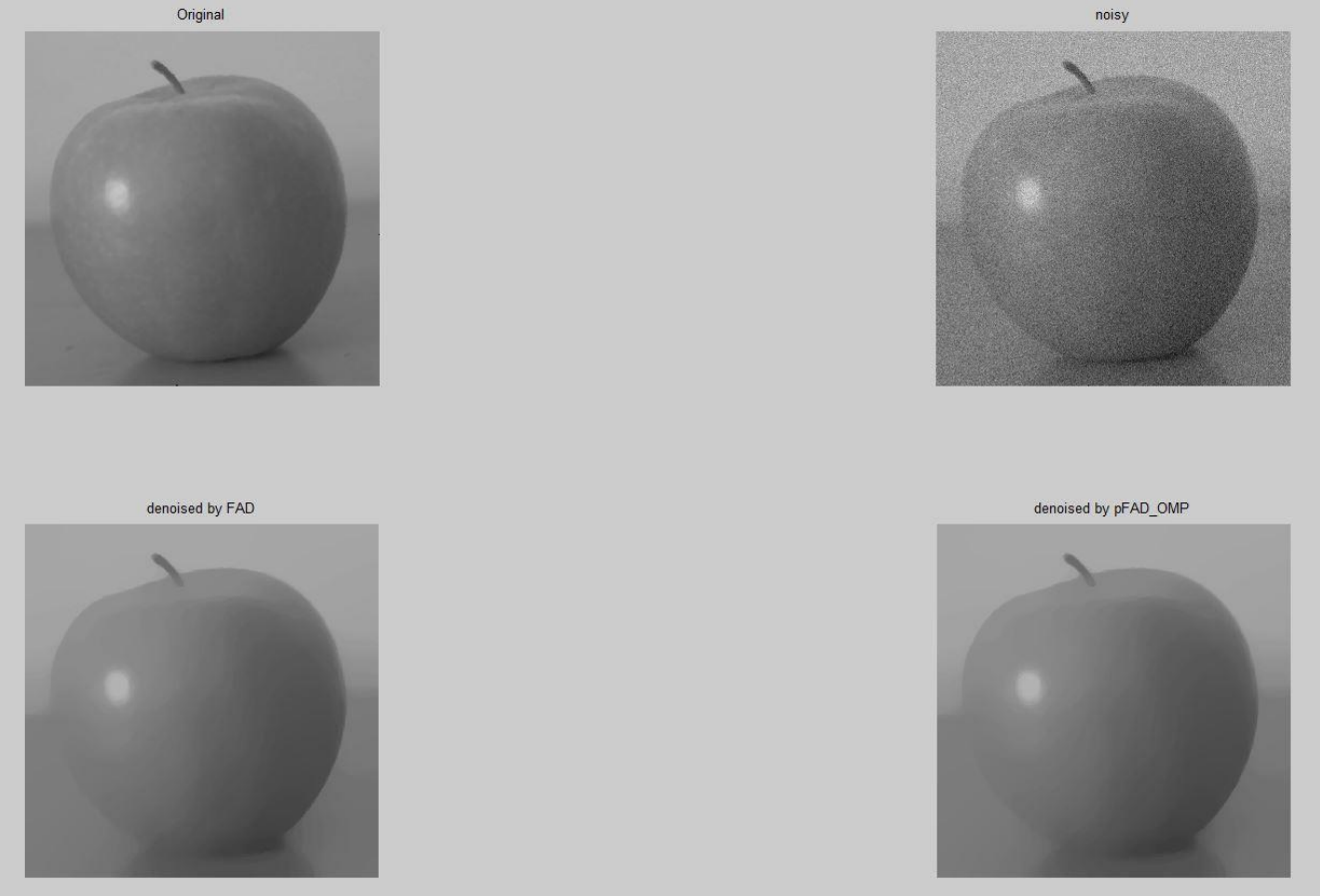


Figure 1. Illustration of the decomposition.

$$\min_{Z,X1,X2,X3} \frac{1}{2}\|Z - Y\|_F^2 + \lambda\|X_k\|_{TV}$$

$$s.t. \quad X_k = Z$$

Matlab Result



Single Channel:
Matlab Serial: 11390.7 ms
Matlab Parallel: 5188.9 ms

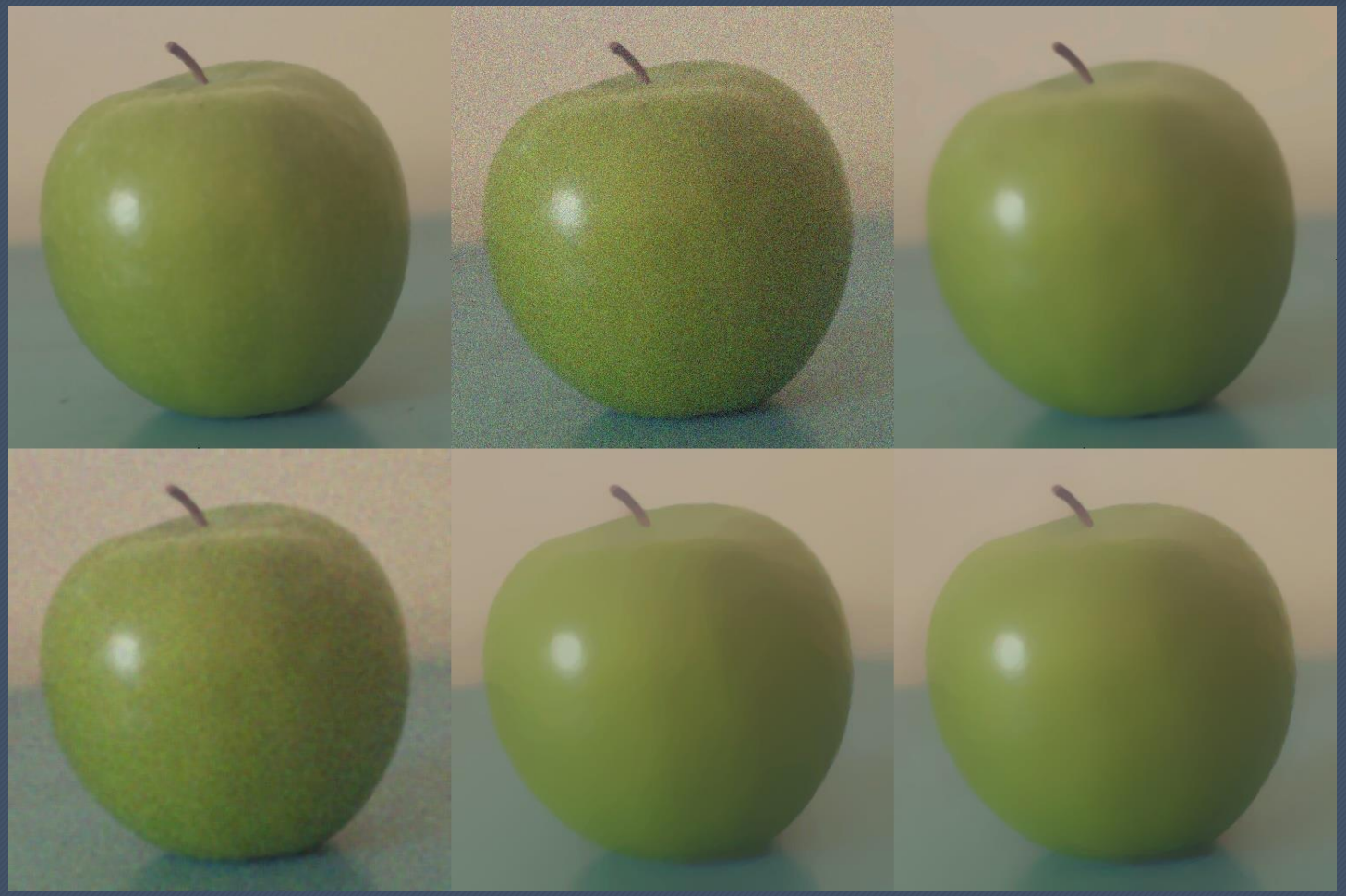
Our Step

- Pure C++
- OpenMP
- CUDA
- Optimization

- Serial : 34454.399 ms
- OpenMP: 17363.747 ms (i5 four cores)
- CUDA : 1758.133 ms (GT540M)
- Opt : 1184.665 ms (GT540M)

Result

Result



Optimization

- double to float: 1.26s → 1.01s
- using double buffer: 1.01s → 933ms
- using for newton and device copy: 933ms → 742ms

Limitation

Time(%)	Time	Calls	Avg	Min	Max	Name
74.59%	742.07ms	153	4.8501ms	3.4379ms	7.8290ms	inside_update(float*, unsigned int*, int, int, float)
8.20%	81.600ms	153	533.33us	528.91us	538.04us	update_sol_u(float*, float*, float*, float*, float*, float*, int, float)
7.94%	78.990ms	459	172.09us	170.88us	173.29us	update_X(float*, float*, float*, float, int, int)
4.50%	44.744ms	6	7.4573ms	7.4131ms	7.4918ms	atomic_reduction_dual(float*, float*, float*, float*, float*, float*, float*, int)
2.89%	28.788ms	3	9.5961ms	9.5760ms	9.6202ms	atomic_reduction_pri(float*, float*, float*, float*, float*, int)
1.30%	12.936ms	150	86.239us	85.533us	88.093us	[CUDA memcpy DtoD]
0.19%	1.9007ms	153	12.423us	11.865us	13.096us	height_boundary_update(float*, unsigned int*, int, int, float)
0.15%	1.5178ms	9	168.64us	166.46us	172.28us	[CUDA memcpy HtoD]
0.10%	991.62us	30	33.053us	1.1840us	105.69us	[CUDA memset]
0.05%	531.38us	24	22.140us	1.4720us	164.22us	[CUDA memcpy DtoH]
0.04%	419.11us	153	2.7390us	2.6110us	3.0080us	width_boundary_update(float*, unsigned int*, int, int, float)
0.03%	314.27us	3	104.76us	104.14us	105.46us	fill_BlkInd(unsigned int*, int, int)

inside_update, because of newton iteration.
Memory pattern.

Division of Labor

赵申剑	Interior update, Tests	34%
陈蕾宇	Boundary update, Tests	33%
徐文康	Stop condition, Tests	32%

Thanks!