



木曾超広視野高速 CMOS カメラ Tomo-e Gozen を支える解析システムの開発

大澤亮 and Tomo-e Gozen Project Members
東京大学理学系研究科 天文学教育研究センター

Outline

Tomo-e Gozen: Next-generation wide-field camera

1. Tomo-e 計画と動画ビッグデータ
2. 超広視野高速カメラを支えるバックエンド
 - 高速化
 - 自動化
 - データ圧縮
3. まとめ

次世代の広視野サーベイ観測

より深く

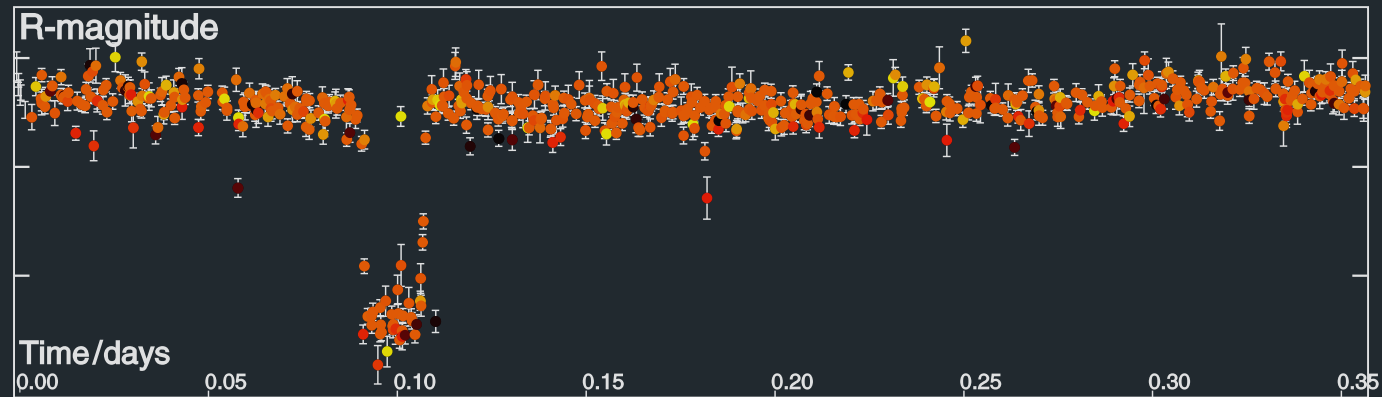
HSC など
10m 級望遠鏡の集光力を活かしたサーベイ観測

より広く

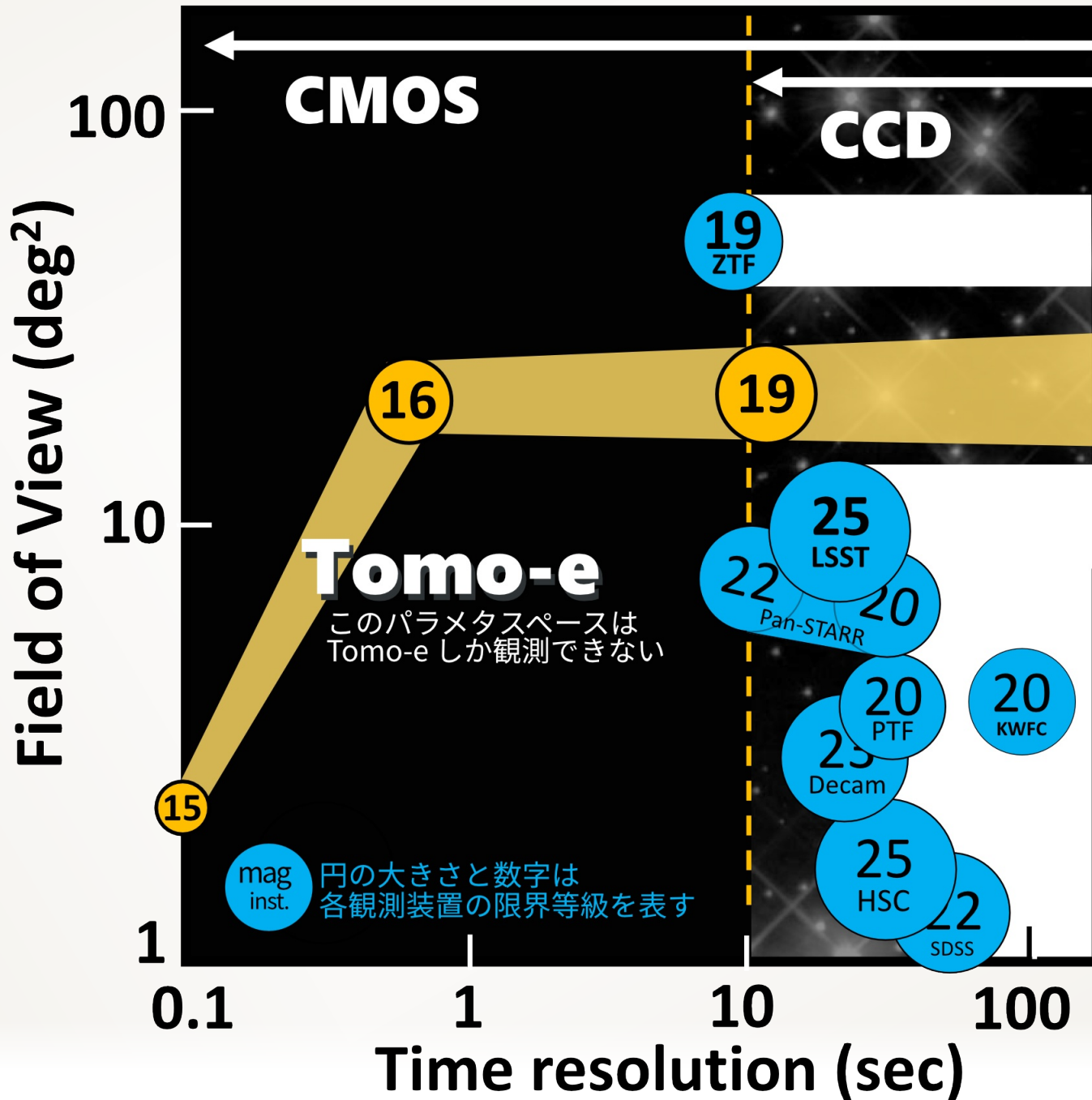
Pan-STARRS など
全天を掃いて基盤となるデータベースを創出する

より速く

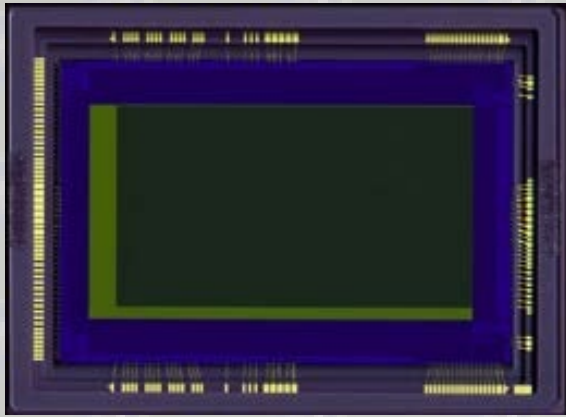
KISS, KISOGP, PTF など
まだ開拓の余地がある領域



広視野サーベイ観測のパラメタスペース



Tomo-e Gozen カメラのデータレート



Canon
35mm Full HD CMOS sensor
2000 × 1128 pixels

$84\text{chip} \times 2000 \times 1128\text{pix} \sim 190\text{Mpix}$
 $190\text{Mpix} \times 16\text{bit} \sim 380\text{MB/frame}$
 $380\text{MB/frame} \times 2\text{fps} \sim \mathbf{760\text{MB/s}}$

参考 Ethernet 接続

100baseT $\sim 16\text{MB/s}$

1000baseT $\sim 160\text{MB/s}$

超広視野高速カメラ

Tomo-e のための解析バックエンド

高速化

自動化

データ圧縮

高速化

基本に忠実にボトルネックを潰す

disk IO の削減 / 計算の効率化 / 並行処理

(⇔ メンテナンス性)

例) KISS パイプラインの高速化

File IO-based reduction から on-memory reduction への変更

1フレームあたりの処理時間	KISS/KWFC 現行パイプライン	Tomo-e 用 プロトタイプ
bias/flat などの基礎処理^①	10.6 秒	0.249 秒
バックグラウンドの推定	1.4 秒	0.146 秒
PSF サイズ (FWHM) の推定	4.5 秒	0.467 秒
アストロメトリ (WCS 情報更新)^{②③}	11.9 秒	0.287 秒
バックグラウンド除去	2.5 秒	0.443 秒
ゼロ等級 (SN=5 等級) 推定^③	9.3 秒	0.0004 秒
合計解析時間	40.1 秒	1.614 秒

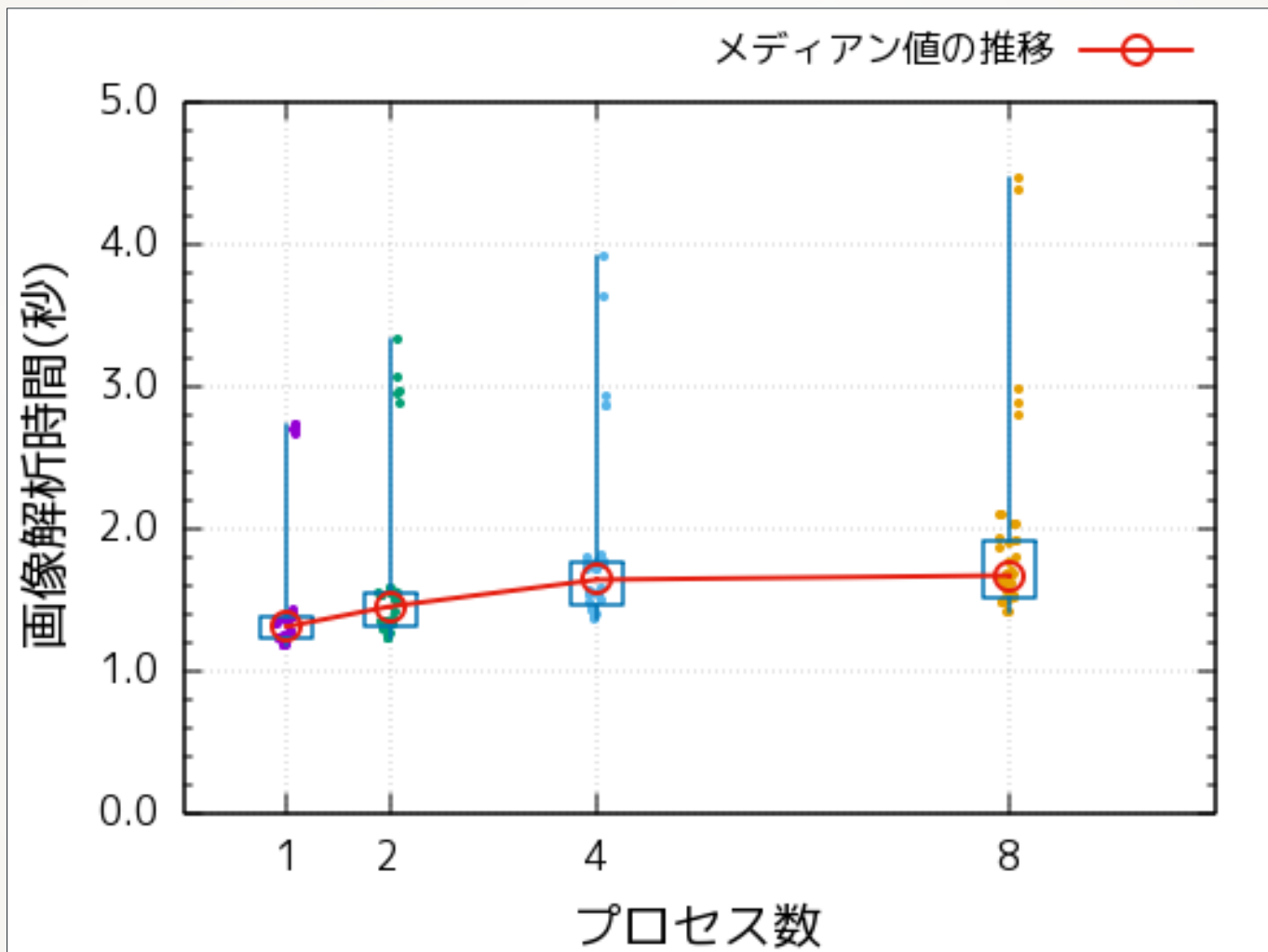
① ファイル IO を減らしたことによるオーバーヘッドの減少

② コストの高い計算を避けることによる計算効率の向上

③ 他のステップで既に計算した内容を流用することによる冗長性の排除

高速化

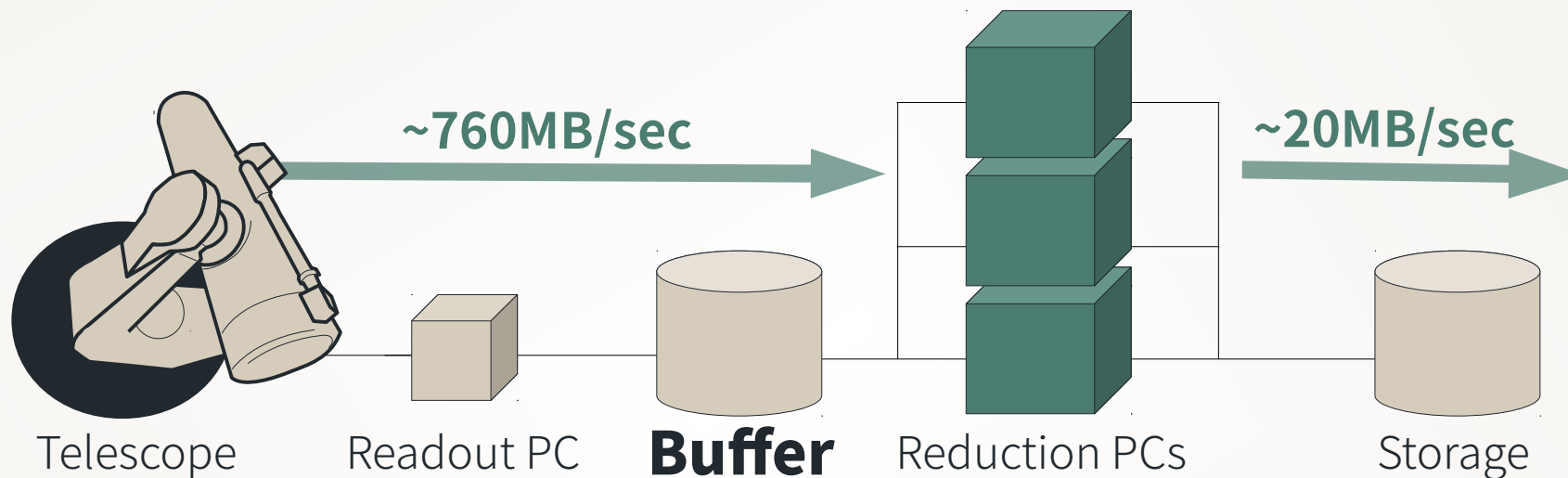
データの並行処理による高速化



画像によって処理時間にかなりばらつきがある

高速化

安定したデータ処理を実現するハードウェア構成



現在進行中

処理速度の違いを緩和するためのバッファの開発

- 10GbE を利用した高速ネットワーク
- SSD RAID による高速ストレージによるバッファの構成

自動化

時間変動減少の効率的な検出

サイエンスターゲット

shock breakout / TNO transit / super flare / etc...

どこに出るかわからない突発的なシグナルを検知する
テンプレート画像との差分処理が広く用いられている

例) iPTF による GRB after glow の探索 (Singer+ 2015)

Table 1
Number of Optical Transient Candidates Surviving Each Vetting Stage

GRB	S/N > 5	RB2 > 0.1	Not Stellar	Not in MPC ^a	Detected Twice	Saved for Follow-up
130702A	14 629	2 388	1 346	1 323	417	11
131011A	21 308	8 652	4 344	4 197	434	23
131231A	9 843	2 503	1 776	1 543	1 265	10
140508A	48 747	22 673	9 970	9 969	619	42
140606B	68 628	26 070	11 063	11 063	1 449	28
140620A	152 224	50 930	17 872	17 872	1 904	34
140623A	71 219	29 434	26 279	26 279	442	23
140808A	19 853	4 804	2 349	2 349	79	12
Median reduction		36%	17%	16%	1.7%	0.068%

~70deg²でこの規模のシグナルが検出される

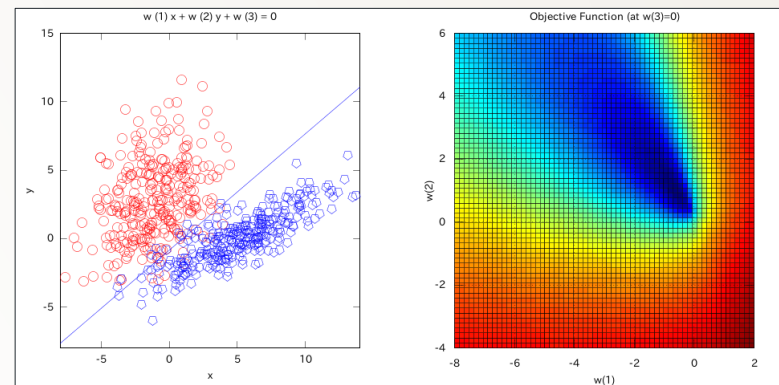
自動化

機械学習による変動信号の分類

目標

信号の特徴量から変動信号を分類する

PTF では既に実用化 / まもなく HSC でも実用化の予定 (?)



イメージ図

座標
FWHM
離心率
etc...

機械学習による分類器

ノイズ
超新星
流れ星
etc...

学習用のデータが必要

KISS/KISOGP 等のデータを利用して分類器のプロトタイプを作成 → Tomo-e につなげる
HSC のチームと情報を共有してやっていきたい

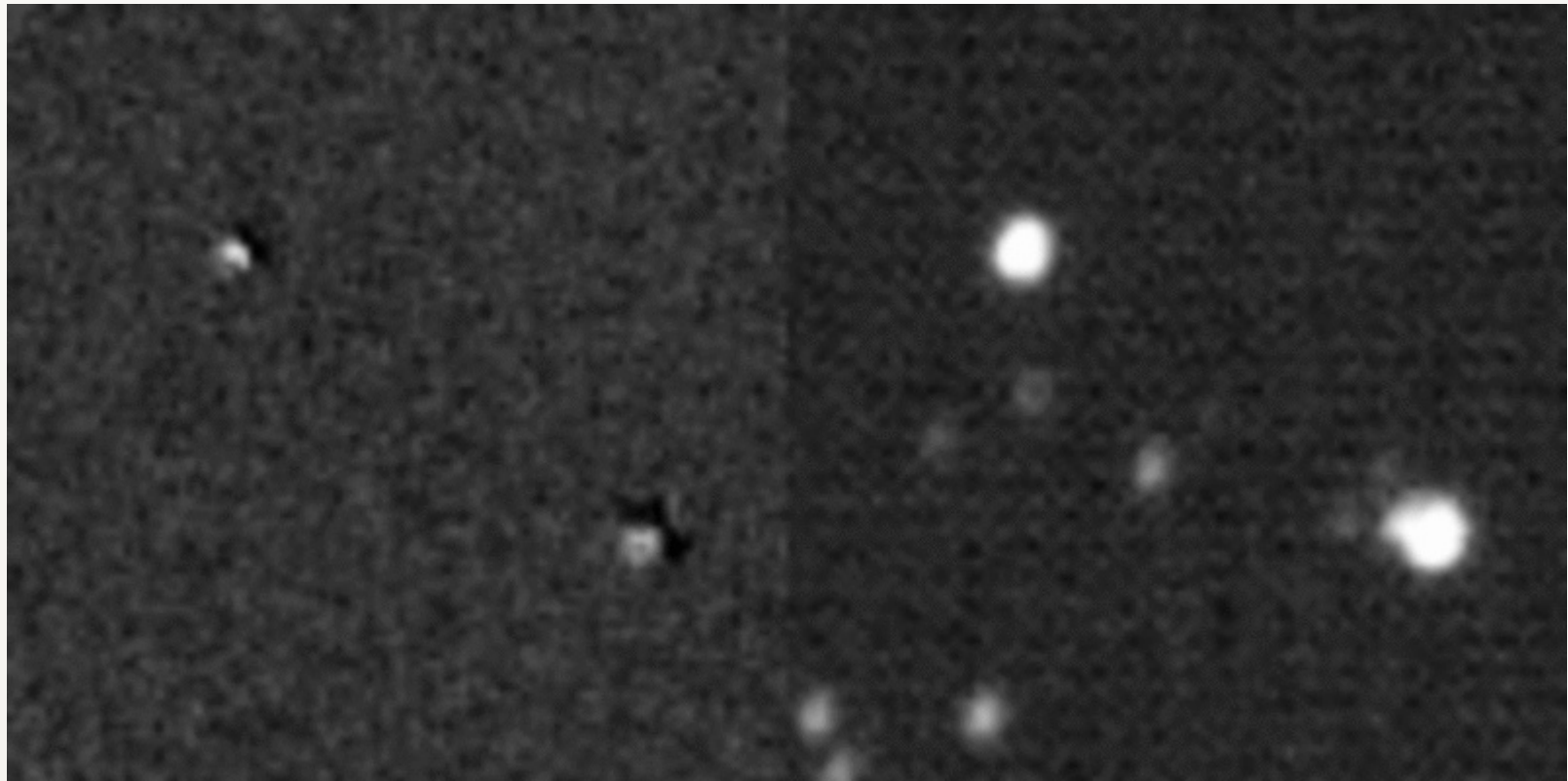
自動化

高速イメージングの世界 (~ 1fps)

1秒スケールでも星のダンスが見える

望遠鏡の振動？シーイングの低周波成分？

追尾エラーや薄雲の通過が差分法で“信号”として検出される

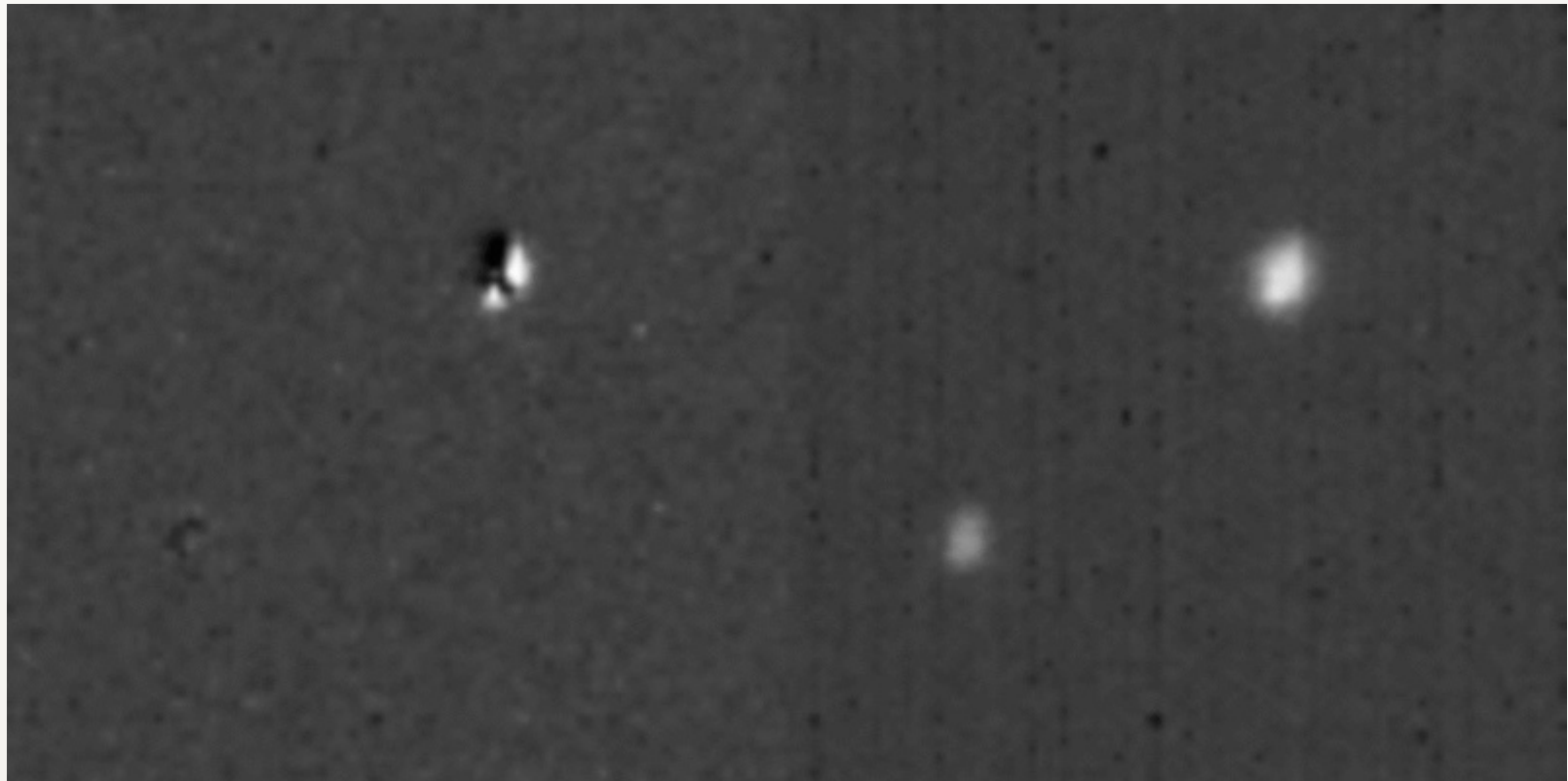


自動化

高速イメージングの世界 (~ 30fps)

PSF がアメーバのように変形する

積分時間が少ないため PSF が十分に収束していない (測光精度にも影響)
これまで時間積分で消えていた不定性が見えてくる？



データ圧縮

必要性

データレート最大で一晩観測をした場合

2Hz 観測 (760MB/s) × 12 時間 ~ **30TB/day**
8TB HDD 4 台

ネットワークで転送すると

30TB / 100Mbps ~ **30days**

cf. 30 TB/day ~ photo uploads in Facebook in 2012

(左) 2012 年頃に建てられた Facebook のデータセンター, (右) 木曾観測所



データ圧縮

サイエンスに必要な十分なデータ保存

全データを保管運用するためにはインフラ面の大工事が必要

→ そんなお金はない

→ 全データを保存する需要？ 1m 望遠鏡でシーイング 4"
他の(強力な)望遠鏡がカバーしている
パラメタスペースで勝負する必要はない

サイエンスに**必要十分なデータ**だけを残す

例) High Cadence (~15min.) 超新星サーベイ

信号が受からなかった領域のデータは時間分解能を落として保存する

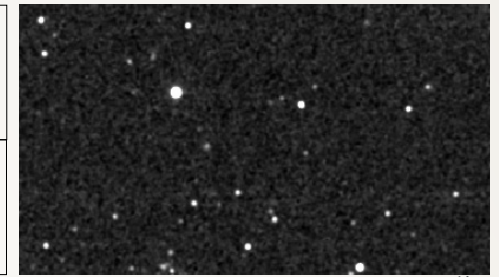
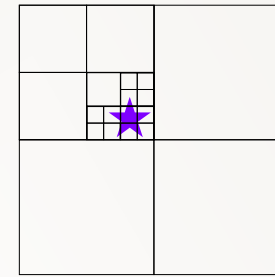
例) TNO 掩蔽モニタリング観測 (30fps)

最低限必要なデータは光度曲線と誤差のみ (画像データはあくまで光度曲線の補佐)

目標となる圧縮率は **1/20 ~1/100** 倍程度

データ圧縮

領域選択的なデータ圧縮手法

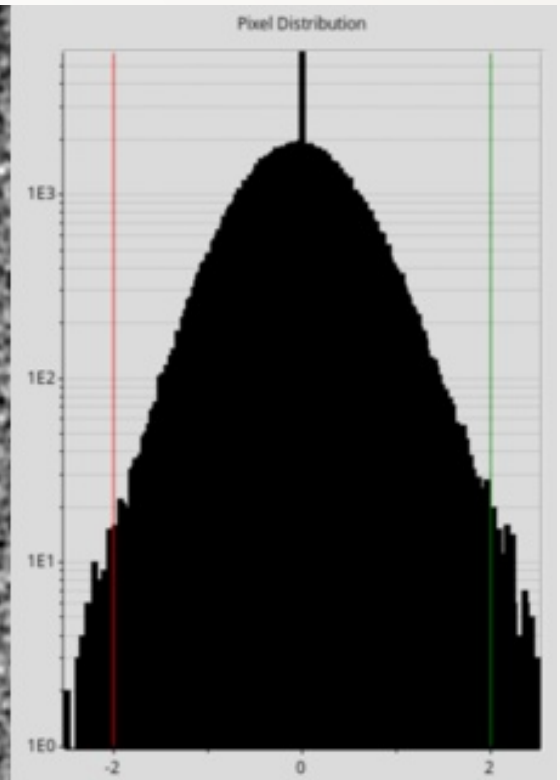
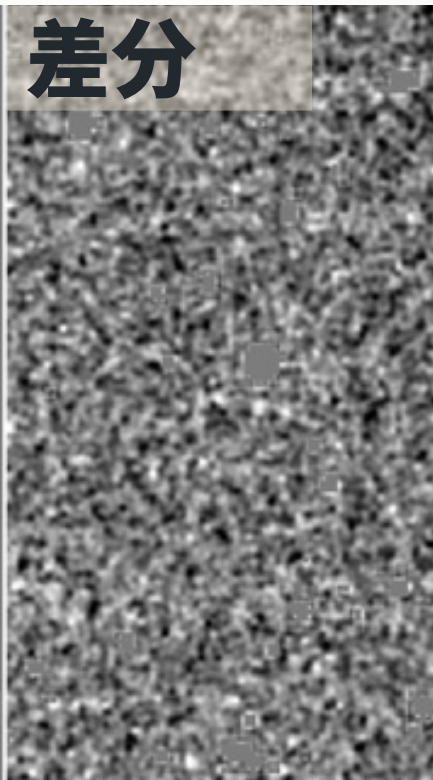
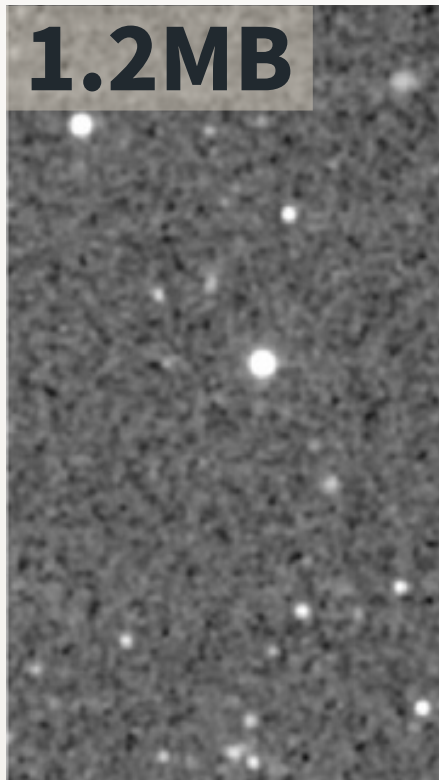


オリジナル画像

Preliminary

データ圧縮方法のアイディアのひとつ

信号のある領域とその周辺以外は代表値で置き換えてしまう
このサンプルではデータを 1/40 程度に圧縮することに成功した



まとめ

1m 望遠鏡 + シーイング 4" でこの先生きのこるには

観測においてユニークなパラメタスペースを攻める

シュミット望遠鏡 + CMOS センサを活かした超広視野高速観測

Tomo-e による天文動画サーベイ

大規模データに対応したバックエンド解析システムの開発

- 高速化 (on memory 処理 / 並行分散処理)
- 自動化 (機械学習による分類器 / 高速撮像観測のノウハウ)
- データ圧縮 (必要十分なデータ量の選択 / 情報の粒度を操作する)

観測計画の立案とパイプラインの設計は不可分

- データを管理するコストが大規模データの生成コストを上回りつつある
- 観測計画と解析システム (特にソフトウェア) の開発は深く連携する必要がある
- おもしろいアイディア・技術に心当たりのある方は大澤にご連絡ください