

Tomo-e Gozenによるサーベイ計画とデータプロダクト

東京大学理学系研究科
天文学教育研究センター 研究員
大澤 亮

theme:

トモエゴゼンのデータが出てくるまで&データの利用方法

table of contents:

1. トモエゴゼン計画とデータの特殊性
2. トモエゴゼンのデータができるまで
3. トモエゴゼンの使い方
4. 進捗状況について

トモエゴゼン計画とデータ

トモエゴゼンの立ち位置について

プロポーザル型

木曾観測所(~2016), 京大 3.8m, すばる, など
KISS/KWFC, KISOGP/KWFC
HSC SSP など

← **トモエゴゼン**

プロジェクト型の特化性能
+
木曾観測所の柔軟さ

SDSS, Pan-STARRS, CRTS, LSST, など

プロジェクト型

トモエゴゼンの特殊性

1. 広い視野 ($\sim 20 \text{deg}^2$) と高い感度を両立
2. オーバヘッドなしの連続読み出しを実現

⇒ これらの特徴を活かした観測を実施したい

トモエゴゼンの特殊性

最大レートで観測し続けた場合

$84 \text{ chip} \times 2.4 \text{ Mpix} \times 4 \text{ byte} \times 2 \text{ fps} \times 10 \text{ 時間} = \text{およそ } 50 \text{ TiB/day}$
($\sim 1600 \text{ MiB/s}$)

6TB HDDに保存した場合: 年間100晩でおよそ3,000万円必要
木曾 ⇄ 外部の通信(フレッツ光)速度: $\sim 125 \text{ MiB/s} \ll \sim 1600 \text{ MiB/s}$

1. 運用可能な状態での保管は困難
2. 木曾観測所から脱出できない

⇒ 観測所内部でサイエンスに必要な情報情報を圧縮だけ抽出

オンサイト解析の必要性

基本方針:

解析済みのデータのみを残す

研究の主軸はタイムドメイン天文学

- ⇒ 時間変動していない信号については重要度が低い
- ⇒ 全動画データを残さなくてもサイエンスが可能

自動解析パイプラインの整備

- ⇒ 多様な観測提案に柔軟に対応したい
- ⇒ 基本となるデータセットを定義・提供する*

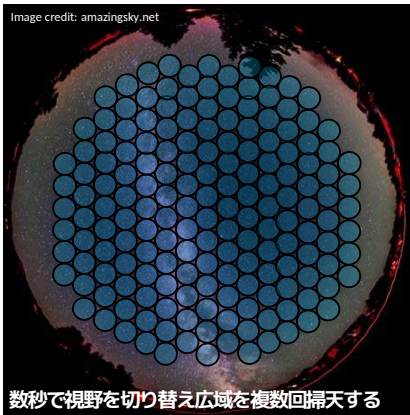
※ 基本データセット以外を用いる観測には共同研究として別途対応する

想定される観測提案の例

広視野高速サーベイ

例: 超新星サーベイ

Image credit: amazingsky.net



数秒で視野を切り替え広域を複数回掃天する

2Hz 定点モニタリング

例: 京大 MU レーダ Tomo-e Gozen 連携流星観測

Image credit: amazingsky.net

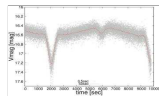


ひとつの視野を 2Hz で長時間モニタリング

提供されるデータ(案)

1. 全天体測光値テーブル at 2 Hz

2Hz で映る全天体 (~18mag.) の測光結果を保存
変動していない天体も含めて全情報を保存する



2. stacked deep images

1 visit での最大積分画像 (3s — 180s) を保存する



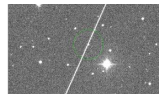
3. cropped movies of fast transients

1, 2 の情報を元に変動があった領域のみ動画情報を残す



4. 高速移動天体スナップショット

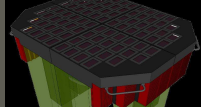
流星や低軌道デブリ, NEOs など 2Hz では線状に現れる信号



トモエゴゼンのデータフロー

木曾シュミット望遠鏡ドーム

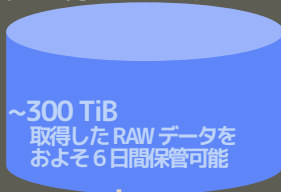
Tomo-eGozen カメラ



~50 TiB/夜

※簡易解析によって16bit→32bit

ドーム内バッファストレージ



ドーム内解析用計算機群

解析によってデータ総量を
およそ~2 TiB/夜まで削減する

現在は4種類のパイプラインを検討中

※中間生成データは残さない

1
全測光値
@2Hz

2
長時間
積分画像

3
抜粋動画
@2Hz

4
高速移動
天体

カメラから出力されたデータ



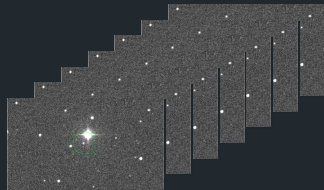
簡易解析(バイアス・フラット処理)

トリミングや32bit化を実施
データレートはおよそ~50TiB/夜になる



ドーム内のストレージに一時保管

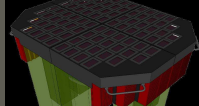
解析処理のバッファとして使用
およそ6日前のデータまで後から参照可能



3次元 FITS データ

木曾シュミット望遠鏡ドーム

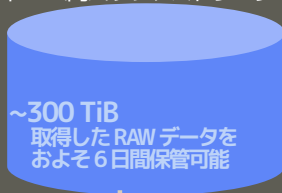
Tomo-eGozen カメラ



~50 TiB/夜

※簡易解析によって16bit→32bit

ドーム内バッファストレージ



ドーム内解析用計算機群

解析によってデータ総量を
およそ~2 TiB/夜まで削減する
現在は4種類のパイプラインを検討中

※中間生成データは残さない



カメラから出力されたデータ



簡易解析(バイアス・フラット処理)

トリミングや32bit化を実施
データレートはおよそ~50TiB/夜になる



ドーム内のストレージに一時保管

解析処理のバッファとして使用
およそ6日前のデータまで後から参照可能



解析用計算機に転送



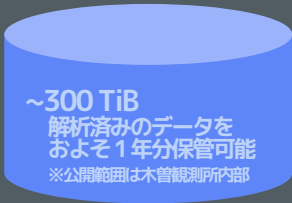
自動解析パイプラインで処理

WCSの貼り付けなどはこの段階で実施予定
必要なデータだけ抽出して~1-2 TiB/夜まで圧縮
基本的に中間生成データは残さない
生成されるデータについては後述

本館のアーカイブに転送
解析済みのデータをおよそ1年分保持可能

木曾観測所本館

Tomo-e Gozen データアーカイブ



※木曾観測所の計算機室にプロジェクト用のスペースを用意

各プロジェクトの計算機へ



物理メディアを郵送して
データセンターにアーカイブ
※ 頻度は月に1度程度を想定

外部アーカイブシステム(予定)

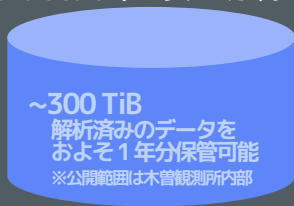
※ 現在 SMOKA チームと協力して検討を進めている

本館のアーカイブに転送
解析済みのデータをおよそ1年分保持可能

木曾観測所本館

※木曾観測所の計算機室にプロジェクト用のスペースを用意

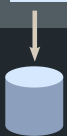
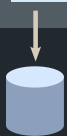
Tomo-e Gozen データアーカイブ



各プロジェクトの計算機へ



物理メディアを郵送して
データセンターにアーカイブ
※ 頻度は月に1度程度を想定



外部アーカイブシステム(予定)

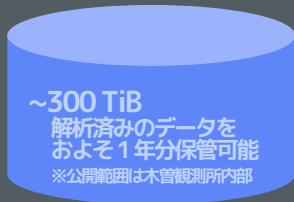
※ 現在 SMOKA チームと協力して検討を進めている

本館のアーカイブに転送
解析済みのデータをおよそ1年分保持可能

木曾観測所本館

※木曾観測所の計算機室にプロジェクト用のスペースを用意

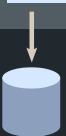
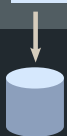
Tomo-e Gozen データアーカイブ



各プロジェクトの計算機へ



物理メディアを郵送して
データセンターにアーカイブ
※ 頻度は月に1度程度を想定



外部アーカイブシステム(予定)

※ 現在 SMOKA チームと協力して検討を進めている

データへのアクセス方法

基本的に観測所内部のアーカイブは外部に非公開

データ転送で観測所への回線を食い潰す恐れがあるため

共同研究プロジェクトにデータのアクセス権

⇒ 本館の計算機室にプロジェクトの計算機を置く

⇒ プロジェクト計算機から本館のアーカイブにアクセス

ドーム内の計算機には直接アクセスさせない方針

※ 外部のアーカイブシステムでのデータ公開も予定 (現在検討中)

データ解析システムの開発について

1. 全天体測光値テーブル at 2 Hz
⇒ Michael システムを参考に構築を進めている★
2. stacked deep images
⇒ clipped mean or mean で足すだけ (ほぼ開発要素無し)
3. cropped movies of fast transients
遅れている (Q1 運用時には間に合わないかも)
4. 高速移動天体スナップショット
流星検出システムの brush up が必要

KWFC(KISS)データベース

```
mysql> show tables;
```

Tables_in_kiss
AGN
IAUSN
KISS_OS
KISS_SN
KISS_Var
KISS_bbs
KISS_phot
LCupload
OSupload
VarS
bcdwarf
blazars
bookmark
check_reduction
check_source
nlsyl
reduction
reduction20160502
score
source
source20160502
users

22 rows in set (0.00 sec)

```
mysql> show fields from source;
```

Field	Type	Null	Key	Default	Extra
proc_id	int(11)	YES	MUL	NULL	
fitsid	varchar(9)	YES		NULL	
num	int(11)	YES		NULL	
x	double	YES		NULL	
y	double	YES		NULL	
RA_deg	double	YES		NULL	
DEC_deg	double	YES		NULL	
FWHM	double	YES		NULL	
elongation	double	YES		NULL	
std_on	double	YES		NULL	
std_off	double	YES		NULL	
fluxmax	double	YES		NULL	
class_star	double	YES		NULL	
flux	double	YES		NULL	
fluxerr	double	YES		NULL	
mag	double	YES		NULL	
magerr	double	YES		NULL	
density	int(11)	YES		NULL	
flag_star	int(11)	YES		NULL	
star_id	bigint(20)	YES		NULL	
flag_gal	int(11)	YES		NULL	
gal_id	bigint(20)	YES		NULL	
source_id	int(11)	NO	PRI	NULL	auto_increment

23 rows in set (0.00 sec)

KISSをベースに Tomo-eらしい情報を付加したい

Optical Flow

- 2 枚の画像から輝度の“移動量”を求めるアルゴリズム (の総称).
- モーショントラッキングやカメラの運動を補正などの応用がある.
- 以下に監視カメラから車の運動ベクトルを計算するサンプルを示す.

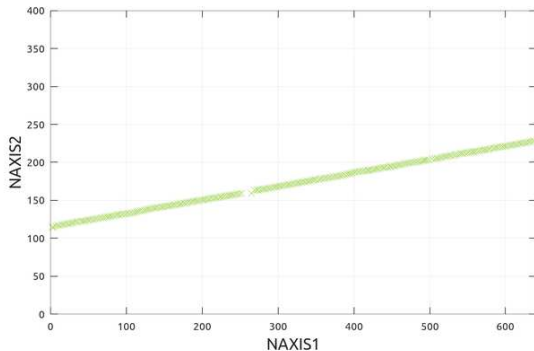
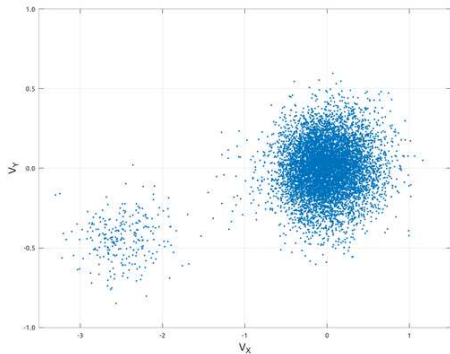


Optical Flow 使ってみた



Optical Flow 使ってみた

- (左): 検出された天体に割り当てられた速度の散布図
- (右): $(V_x, V_y) = (-2.5, -0.5)$ 付近の天体を抜き出したもの



Tomo-e Gozen + Optical Flow

- 視野に依っては Tomo-e Gozen 1 チップの視野で 2000 天体を含むこともある
 - 3分観測すると 70 万件の測光データとなる
 - この中から座標のマッチングで移動天体を抽出するのはなかなか大変
- Tomo-e Gozen は 2 Hz の観測から高速移動天体の速度を推定できる
 - 観測したデータに移動天体が含まれるかを自動判定できる
 - 特にモニタリング観測で効果を発揮
 - 明るい天体ならサーベイ観測でも特定可能
 - 計算量はさほど重くないので自動解析パイプラインに組み込みたい

まとめ

トモエゴゼン計画:

特化型かつ柔軟な観測に対応することを目指す
プロジェクト単位での共同研究を想定

データ運用計画:

全データ保存は難しい ⇒ 重要な情報だけ抽出する
データは木曾でアクセスして木曾で解析する

検討事項:

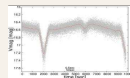
最終的なアーカイブとしてデータのホスティングを依頼
基本パイプラインで残す情報セット
より複雑な観測提案に対する共同研究体勢

議論・検討事項

議論 — 提供されるデータセット

1. 全天体測光値テーブル at 2 Hz

2Hz で映る全天体 (~19 mag.) の測光結果を保存
変動していない天体も含めて全情報を保存する



2. stacked deep images

1 visit での最大積分画像 (3s — 180s) を保存する



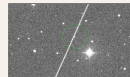
3. cropped movies of fast transients

1, 2 の情報を元に変動があった領域のみ動画情報を残す



4. 高速移動天体スナップショット

流星や低軌道デブリ, NEOs など 2Hz では線状に現れる信号



議論 — より柔軟なプロジェクトに

特殊な観測計画の実施について

高速部分読み出し (e.g., TNO survey, pulsar survey)
non-sidereal objects (e.g., debris, meteors)
対物グリズム分光観測

⇒ 解析パイプラインの開発も含めた共同研究として考えたい

動画をより深く解析する計画について

動画解析技術の応用, 科学用途指向動画圧縮 (e.g., GoDec 圧縮)
ハードウェアパイプラインによる高速動画処理 (e.g., 重ねあわせ法)


⇒ 共同研究計算機をドーム内に増設することを検討

提供されるデータ(量)について

- 1. 全天体測光値テーブル at 2 Hz** **monitoring: ~200GB**
survey: ~30GB
2Hz で映る全天体 (~18mag.) およそ 10^5 天体の測光結果を保存
測光結果リストを text テーブル or データベースとして提供
- 2. stacked deep images** **monitoring: ~75GB**
survey: ~1.5TB
サーベイなら 3s, モニタリングなら 180s 積分画像に相当
- 3. cropped movies of fast transients** **monitoring: ~250GB**
survey: ~40GB
10×10 の領域を 8,000 天体分確保, 3次元 FITS 画像として提供
- 4. 高速移動天体スナップショット** **monitoring: ~500GB**
survey: ~150GB
流星であれば 10^3 – 10^4 イベントが期待, 3次元 FITS 画像として提供

LSSTのデータ計画

LSST From the User's Perspective



- A stream of ~10 million time-domain events per night, detected and transmitted to event distribution networks within 60 seconds of observation.
- A catalog of orbits for ~6 million bodies in the Solar System.

Level 1

- A catalog of ~37 billion objects (20B galaxies, 17B stars), ~7 trillion observations ("sources"), and ~30 trillion measurements ("forced sources"), produced annually, accessible through online databases.
- Deep co-added images.

Level 2

- Services and computing resources at the Data Access Centers to enable user-specified custom processing and analysis.
- Software and APIs enabling development of analysis codes.

Level 3