

RAPORT

z przebiegu praktyk studenckich

Aleksander Tyburek

Spis treści:

1. Oprogramowanie do testów wydajności
2. Środowisko pracy
3. Struktura danych
4. Skrypty
5. Procedura testowania wydajności
6. Wyniki
7. Podsumowanie

Temat praktyk: „*Wyznaczenie efektywności wykrywania nowych obiektów na niebie w danych >>Pi of the Sky<<*”

Projekt „Pi of the Sky” ma na celu wykrywanie krótkozmiennych zjawisk zachodzących na niebie, takich jak błyski optyczne towarzyszące błyskom gamma czy wybuchy gwiazd nowych. Pomiaru wykonywane są przy użyciu kamer CCD; dane z nich gromadzone są w bazie oraz poddawane działaniu algorytmów poszukujących w nich przytoczonych zjawisk. Zadaniem algorytmów analizujących poszczególne zdjęcia jest wykrywanie obiektów będących potencjalnymi nowymi, przy jednoczesnym odrzucaniu różnego rodzaju przypadków tła; artefaktów, szumów, kryształków lodu osiadających na obiektywie kamery, sztucznych satelitów itp.

Celem ćwiczenia było przetestowanie wydajności działania tych algorytmów oraz zbadanie jej zależności od różnych parametrów towarzyszących zarówno samym pomiarom jak i analizie danych.

1. Oprogramowanie do testów wydajności

Wydajność algorytmów szukających nowych testowana jest za pomocą odpowiednio przygotowanych zdjęć nieba. Przygotowane do tego celu oprogramowanie wycina z podanych zdjęć sample - obrazy gwiazd - i wkleja je na wybrane klatki w losowo wybieranych pozycjach; ilość wklejonych sampli oraz zakresy ich jasności podawane są jako parametry programu.

Testowanie wydajności polega na uruchamianiu algorytmów szukających gwiazd nowych w przygotowanych zdjęciach z wklejonymi samplami. Znając liczbę wklejonych na poszczególne klatki sampli oraz ich jasność można określić wydajność wykrywania nowych przez algorytm (stosunek liczby wykrytych do wklejonych) oraz zbadać jej zależność od jasności.

2. Środowisko pracy

Oprogramowanie projektu „Pi of the Sky” funkcjonuje w systemie Linux i to właśnie na tej platformie wykonywane były wszelkie prace; dostęp do serwera z oprogramowaniem oraz danymi projektu możliwy był poprzez połączenie za pośrednictwem protokołu SSH.

Dane z pomiarów „Pi of the Sky” oraz dane wynikowe działających algorytmów umieszczone zostały w bazie danych, w odpowiednich tabelach. Do pracy wykorzystana została baza danych PostgreSQL.

3. Struktura danych.

Dane pomiarowe zgromadzone są w następującej hierarchii:

data_wykonania → numer kamery → pliki klatek

W użyciu są cztery kamery, z których dane posiadają w nazwach plików - w zależności od numeru - przedrostki cam[nr], [nr] dla kolejnych kamer wynosi 32, 34, 36, 40.

Dla każdej z kamer danej nocy jest od kilku do kilkudziesięciu klatek.

Identyczna hierarchia zastosowana jest do przetworzonych danych służących testom wydajnościowym, tj. klatek z wklejonymi samplami.

Dane związane z pomiarami oraz wklejonymi samplami zgromadzone są w tabelach:

Frame – tabela zawierająca dane charakterystyczne dla pojedynczej klatki; data wykonania zdjęcia, nr klatki, ścieżka dostępu do pliku, rozmiary, położenie obserwowanego obszaru nieba w równonocnym układzie współrzędnych, ustawienie kamery w tej pozycji.

PastedEvents – zawiera dane dotyczące pojedynczego z wklejonych sampli; nr klatki, ścieżka do pliku, współrzędne, jasność itp.

Measurements – zawiera szczegóły pomiarów uzyskanych z wykonanych zdjęć: współrzędne, jasności, niepewności pomiarowe itp.

AstFile – tabela zawierająca dane zawarte w plikach *.ast, uzyskiwanych w wyniku działania procedury astrometrii, porównującej zebrane dane z katalogiem gwiazd TYCHO w celu przekształcenia współrzędnych zdjęcia na współrzędne równonocne.

NovaEvents – tabela zawierająca dane na temat obiektów zidentyfikowanych przez algorytm jako gwiazdy nowe.

4. Skrypty

Do zrealizowania zadania napisane zostały skrypty w powłoce BASH oraz skrypty (funkcje) w języku PL/pgSQL, uruchamiane w systemie zarządzania bazami danych PostgreSQL.

Lista skryptów:

Skrypty BASH:

do_add_samples!
dir_sam2sql
sm2sqlsched

Pliki z definicjami procedur PL/pgSQL:

cfe.sql
nightFramesEfficiency.sql
nEffVsMag.sql
AfVsMag.sql

Opis skryptów:

do_add_samples! [original frames directory] [new frames directory] [number of samples per frame] [min night] [max night]

Skrypt bash-owy służący do zautomatyzowanego wywoływania programów do wycinania i wklejania sampli. W katalogu samples (jeśli taki nie istnieje, tworzy go) umieszcza wycięte sample z wszystkich klatek z nocy należących do przedziału <min night; max night>. Następnie wywołuje program put_samples i w katalogu podanym jako drugi parametr tworzy klatki z wklejonymi samplami, zgodnie z hierarchią opisaną w punkcie 3.

Parametry:

- ścieżka dostępu do katalogu zawierającego katalogi z nocami;
- ścieżka dostępu do katalogu wynikowego;
- liczba sampli wklejanych na pojedynczą klatkę;
- najwcześniejsza noc z przedziału;
- najpóźniejsza noc z przedziału.

Noce podawane są jako liczby utworzone z dat, w formacie yyyymmdd, np. 20110602.

dir_sam2sql [frames with samples directory for a single cam]

Skrypt bash-owy służący do przekonwertowania plików *.sam (utworzonych w wyniku działania programu wklejającego sample) na pliki *.sql i załadowania ich do bazy danych.

Jako parametr przyjmowana jest ścieżka dostępu do katalogu zawierającego pliki *.sam, w których znajdują się dane dotyczące wklejonych sampli. Skrypt dla każdego z nich tworzy plik *.sql, w którym znajdują się wpisy umieszczające te dane w tabeli PastedEvents. Następnie każdy z plików *.sql jest ładowany do bazy scan_nova_eff_1.

sm2sqlsched [night dir]

Skrypt bash-owy służący do zautomatyzowania ładowania plików *.sql do bazy.

Jako parametr przyjmowana jest ścieżka do katalogu nocy, w którym zawarte są podkatalogi z plikami *.sam dla każdej z czterech kamer. Dla każdego z czterech podkatalogów wywoływany jest skrypt **dir_sam2sql**.

cfe.sql – skrypt PL/pgsql, definiujący funkcję **CalcFrameEfficiency(frame name)**

Funkcja CalcFrameEfficiency służy do obliczenia wydajności astrometrii i wykrywania nowych dla danej klatki, której nazwa podawana jest jako parametr. Funkcja ta wywoływana jest w ciele funkcji nightFramesEfficiency.

W pierwszej kolejności liczona jest liczba sampli wklejonych na daną klatkę, zgodnie z danymi umieszczonymi w tabeli PastedEvents. Następnie liczone są:

- liczba obiektów znajdujących się w tabeli AstFile, które znajdują się w danej klatce oraz spełniają warunek wystarczająco małej odległości od innego obiektu z tabeli PastedEvents dla tej samej klatki (odległość liczona jest we współrzędnych w równonocnym układzie współrzędnych);
- liczba obiektów znajdujących się w tabeli Measurements, spełniających analogiczne warunki jak w pierwszym przypadku, oraz zaflagowanych jako „true” w polu „new_star”;
- liczba obiektów znajdujących się w tabeli NovaEvents, spełniających analogiczne warunki jak w pierwszym przypadku;

Efektywności kolejnych etapów algorytmu obliczane są jako stosunki każdej z trzech wyżej wymienionych liczb do liczby sampli wklejonych na daną klatkę. Funkcja wypisuje te wartości na ekran.

nightFramesEfficiency.sql – skrypt PL/pgsql, definiujący funkcję **nightFramesEfficiency(night)**

Funkcja nightFramesEfficiency wywołuje funkcję CalcFrameEfficiency dla wszystkich klatek z każdej z kamer, uzyskanych dla danej nocy. Data podawana jest jako argument w formacie yymmdd (np. dla daty 02.06.2011 argumentem funkcji jest '110602'). Funkcja sprawdza wartość pola errcode każdej z klatek i pomija klatki, dla których jest ona niezerowa.

nEffVsMag.sql – skrypt PL/pgsql, definiujący funkcję **nFffVsMag(night)**

Funkcja służy wyznaczeniu zależności średniej efektywności wykrywania nowych od ich jasności dla danej nocy. Jako argument podawana jest data w formacie yymmdd.

Dla każdej z kamer liczona jest zależność wydajności od jasności wg następującego

schematu:

Wprowadzona jest wielkość `mag_step`, która określa szerokość przedziałów jasności gwiazdowych, dla których określana jest efektywność. Przedziały określone są w zakresie od 7 do 14 magnitudo, szerokość przedziału wynosi 1.0 magnitudo (określona jako stała w kodzie, można ją łatwo zmodyfikować). Dla każdego z przedziałów, w pętli iterującej po wszystkich klatkach danej kamery, obliczane są dwie liczby:

- `mag_pe_count` – liczba obiektów z tabeli `PastedEvents`, które znajdują się w danej klatce oraz ich jasność zawarta jest w danym przedziale;
- `mag_ms_count` – liczba obiektów z tabeli `Measurements`, które znajdują się w danej klatce, zaflagowane są jako „true” w polu „new_star”, oraz znajdują się wystarczająco blisko (w promieniu $r < 100$) analogicznych obiektów z tabeli `PastedEvents` o jasności zawartej w danym przedziale.

Wydajność wyznaczana jest jako ich stosunek:

$$\text{mag_eff} = \text{mag_ms_count} / \text{mag_pe_count}$$

Średnia wydajność otrzymywana jest poprzez zsumowanie wydajności dla każdej z klatek z osobna i podzielenie tej sumy przez ilość klatek.

Funkcja wypisuje na ekran zależność średniej efektywności dla przedziałów jasności w zakresie od 7 do 14 magnitudo.

AfVsMag.sql – skrypt PL/pgsql, definiujący funkcję **AfVsMag(night)**

Funkcja służy wyznaczeniu zależności średniej efektywności wykrywania nowych od ich jasności dla danej nocy. Jako argument podawana jest data w formacie `yymmdd`.

Funkcja działa podobnie jak `nEffVsMag` z tą różnicą, że wydajność liczona jest na podstawie tabeli `AstFile`, a nie `Measurements` – funkcja sprawdza więc, w jaki sposób efektywność astrometrii zależy od jasności wklejonych sampli.

5. Procedura testowania wydajności

Poniżej przedstawiona jest krok po kroku procedura testowania wydajności algorytmów szukających nowych dla pomiarów wykonanych jednej nocy – np. 20110602.

1. Przygotowanie klatek z wklejonymi samplami.

```
cd /data1/results/students/atyburek/samples
./do_add_samples! /data1/results/SCAN_PIPELINE ./frames_with_samples 100
20110602 20110602
```

2. Katalogowanie.

```
cd /data1/results/students/atyburek/samples/frames_with_samples
./do_catalog! 20110602
./loadast_night.sh scan_nova_eff_1 cam
```

3. Uruchomienie algorytmu szukającego nowych.

```
cd /data1/results/students/atyburek/msok_test
./run_nova_algo.sh 20110602
```

4. Załadowanie wklejonych sampli do bazy danych (tabela `PastedEvents`).

```
cd /data1/results/students/atyburek/samples/frames_with_samples
./sm2sqlsched 20110602
```

5. Przeprowadzenie testów wydajności.

```
pidb! scan_nova_eff_1
>select nightFramesEfficiency('110602');
>select nEffVsMag('110602');
>select AfVsMag('110602');
```

6. Wyniki

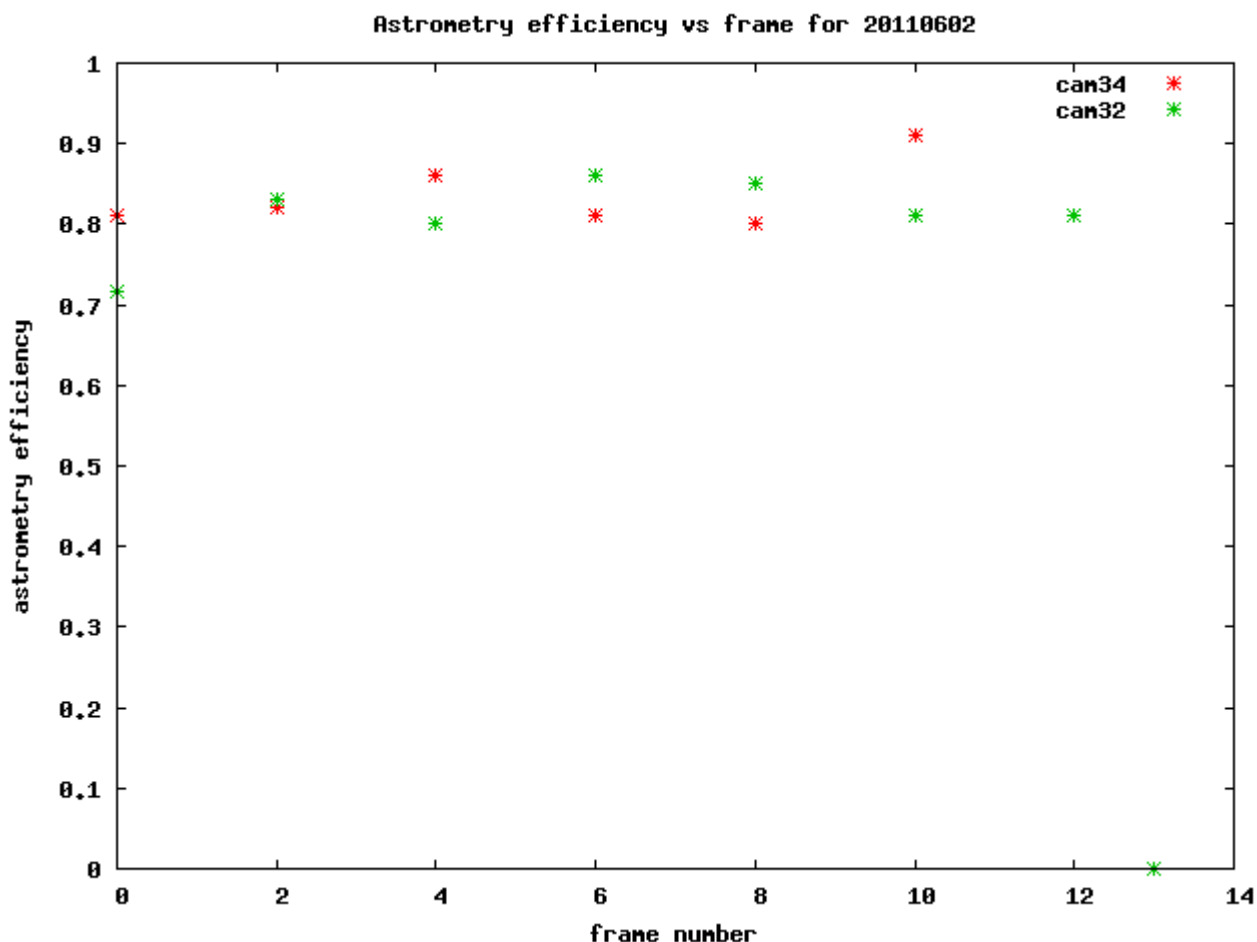
Testy przeprowadzone zostały dla danych zebranych z czterech nocy: 20110602, 20110603, 20110604, 20110608.

Algorytm wykrywania nowych uruchomiony został dwukrotnie dla każdej z nocy, z różnymi wartościami parametru `min_obs_field`.

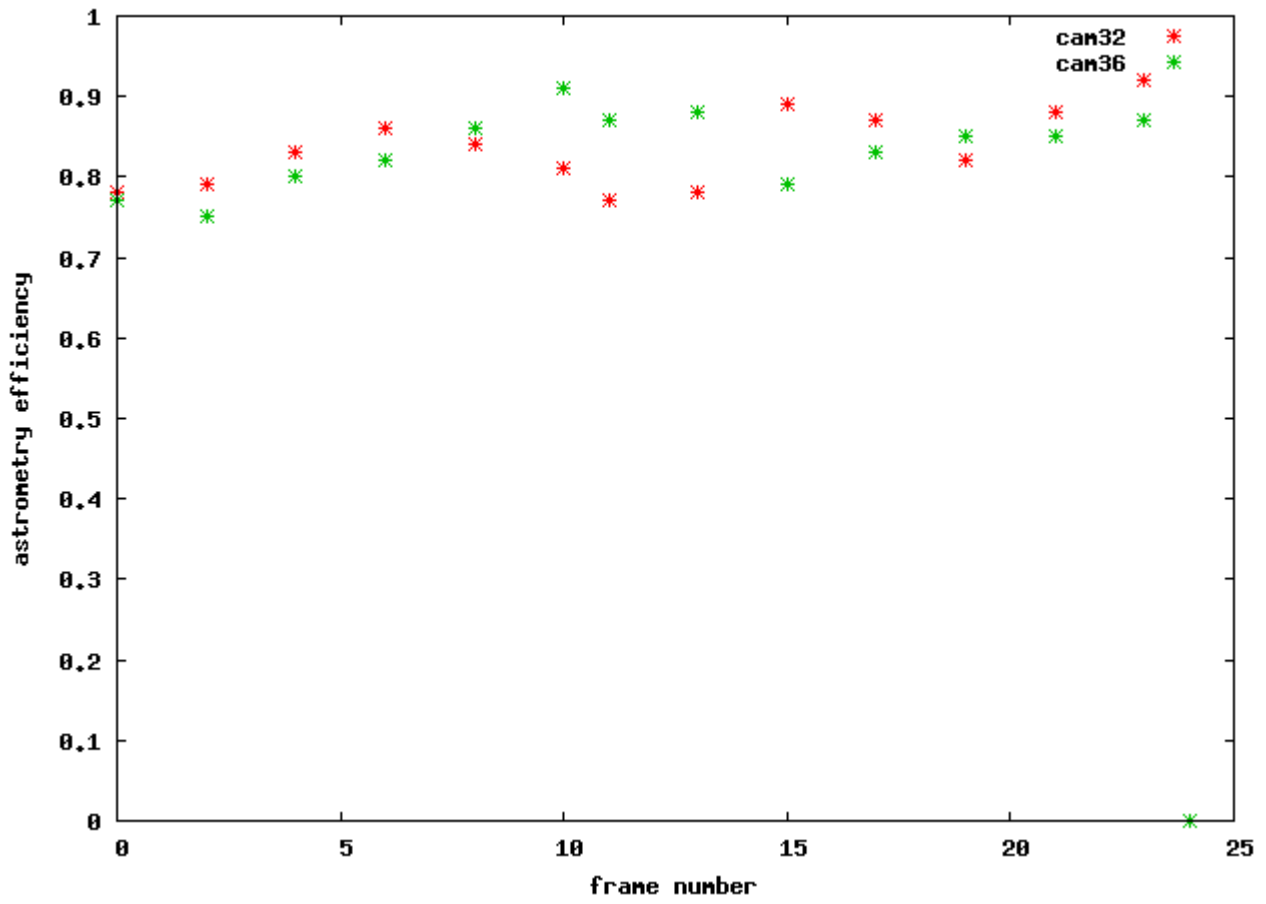
Parametr `min_obs_field` określa minimalną liczbę wykonanych już obserwacji danego obszaru nieba, która wymagana jest do tego, by w tym obszarze algorytm wyszukiwał nowe. Jeśli liczba przeprowadzonych obserwacji tego obszaru jest mniejsza niż `min_obs_field`, żadne nowe nie będą w nim rejestrowane.

Wykonane wykresy przedstawiają przykładowe dane uzyskane przy pomocy sporządzonych skryptów.

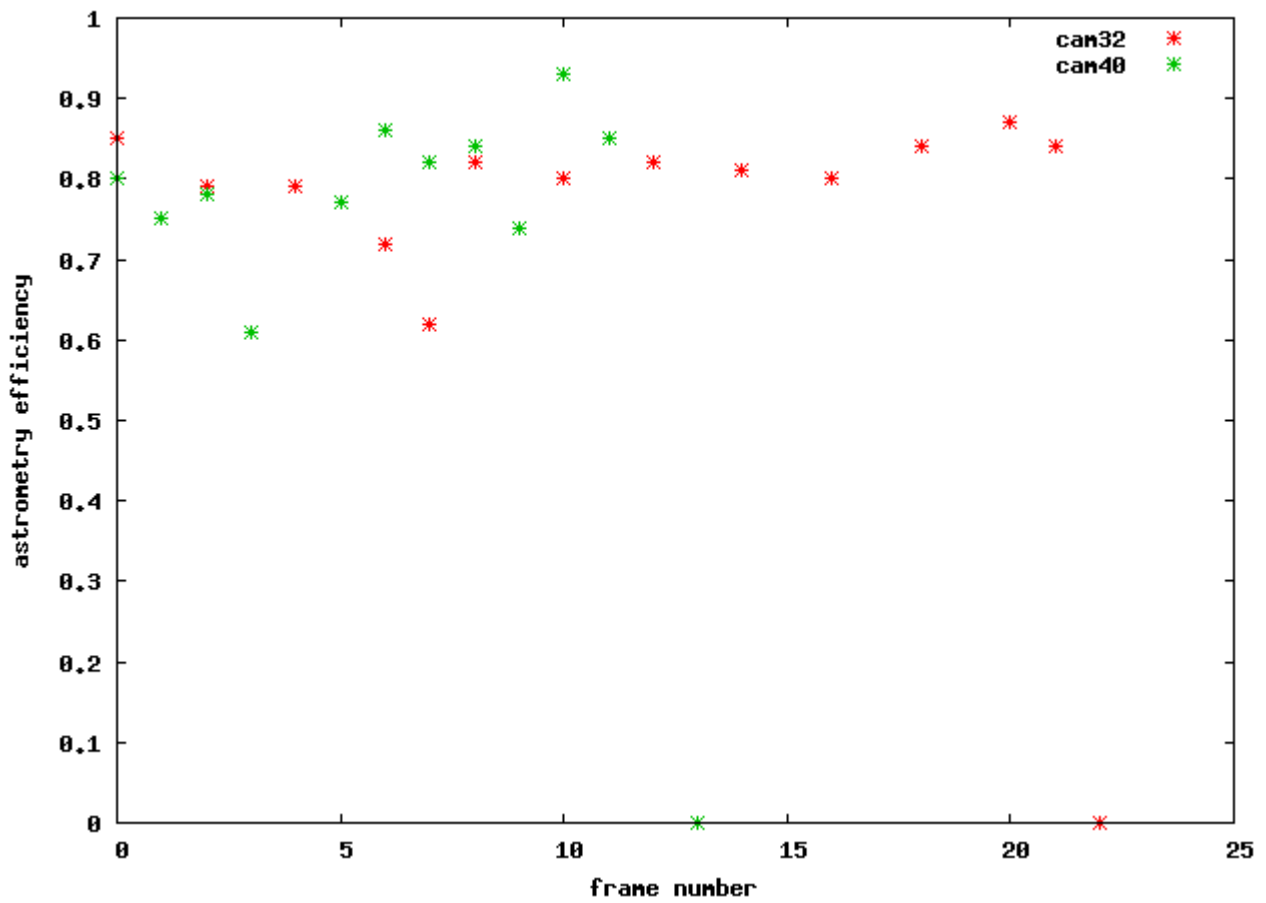
1. Wydajność astrometrii w funkcji klatki dla wybranych kamer:



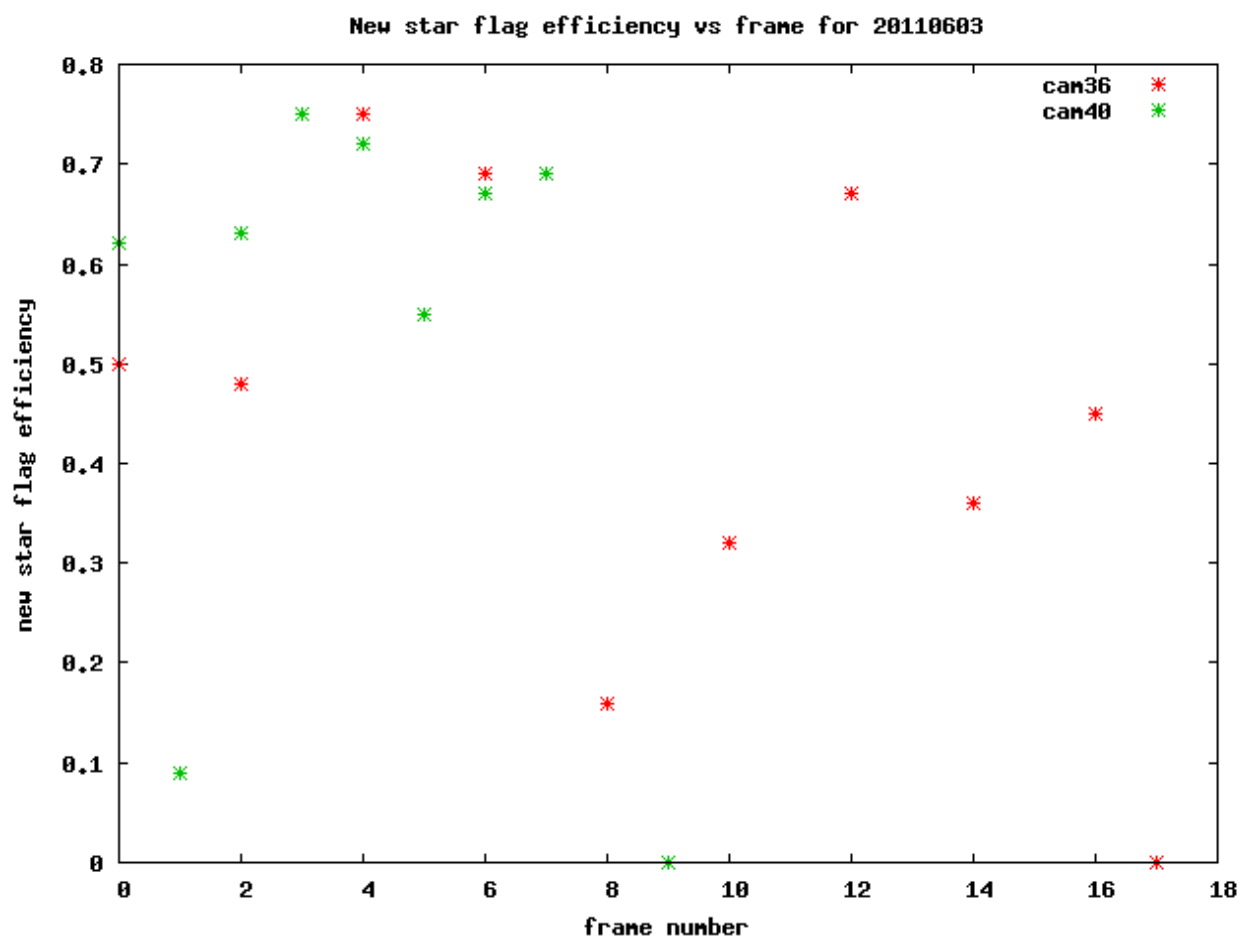
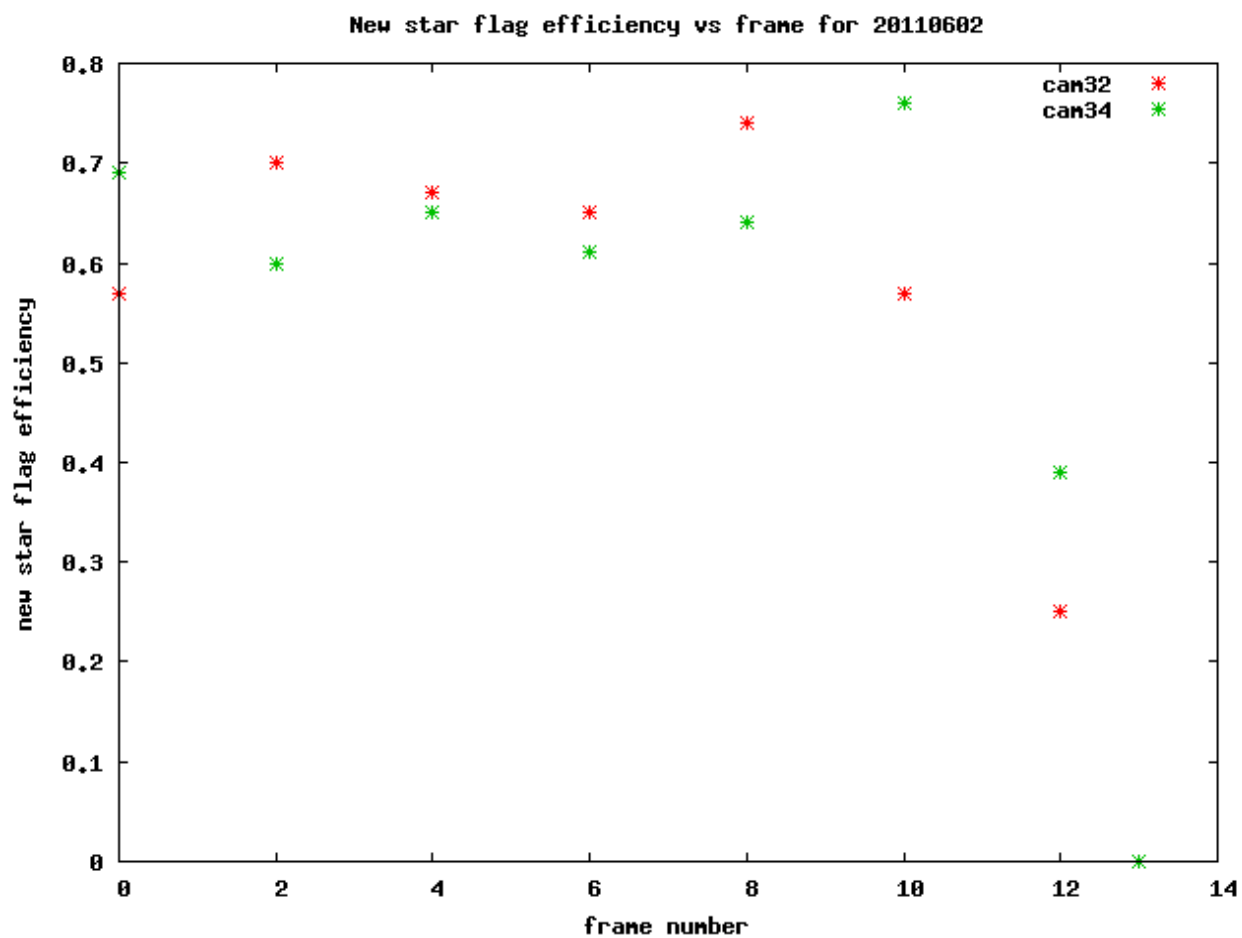
Astrometry efficiency vs frame for 20110604



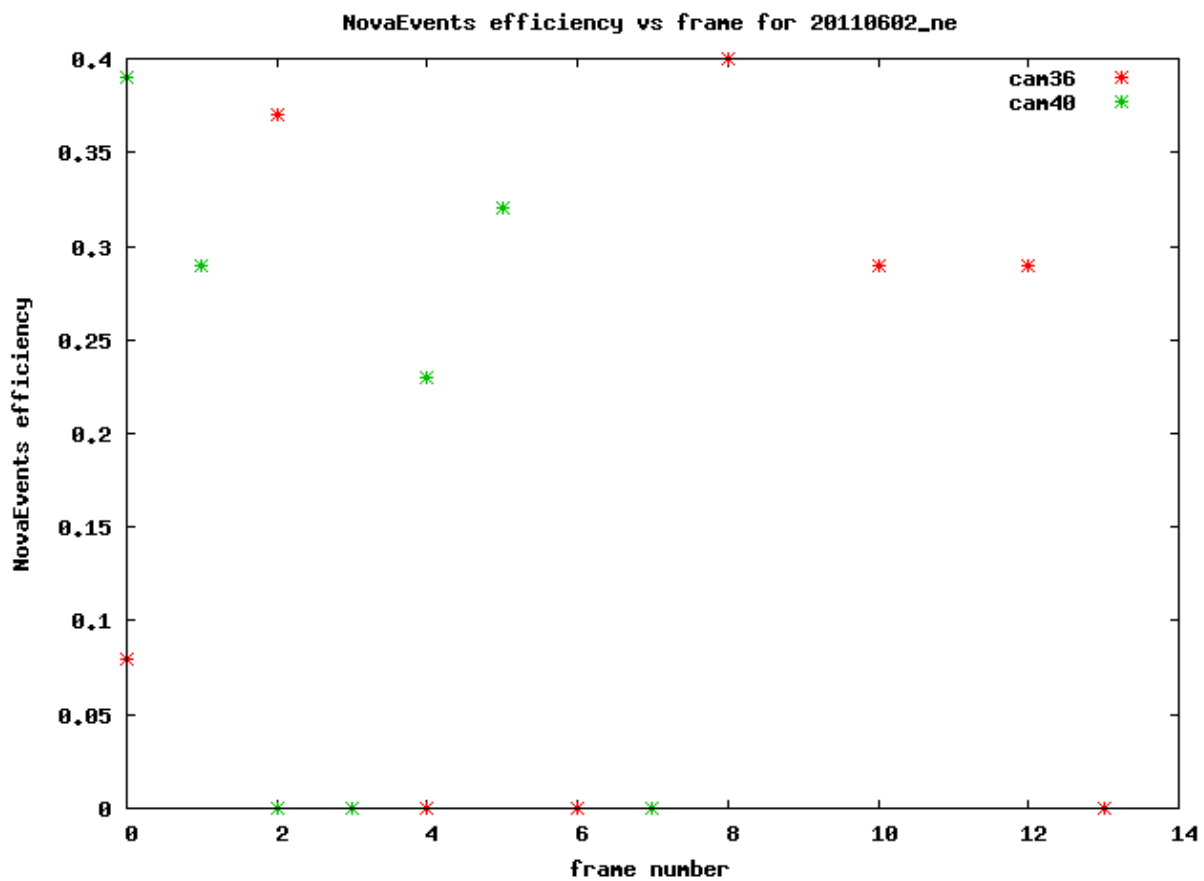
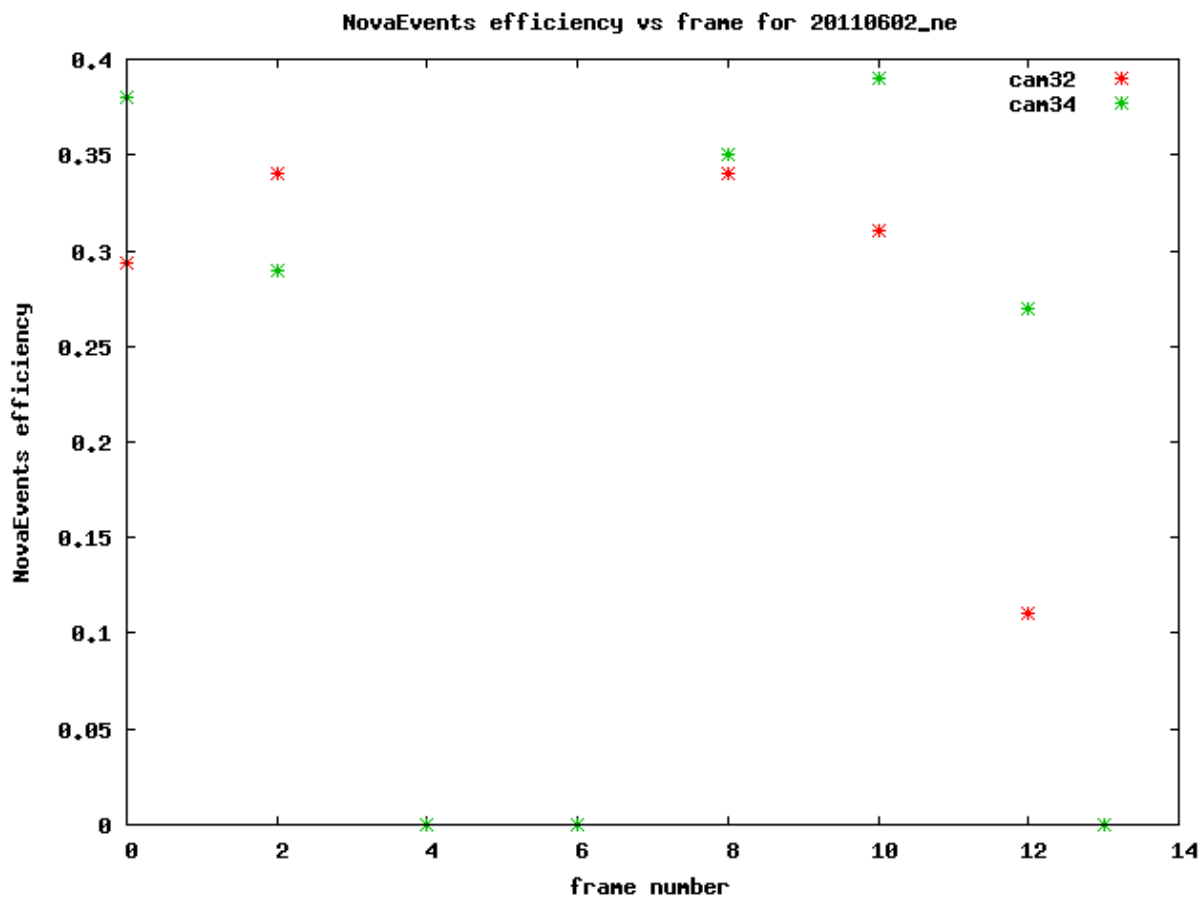
Astrometry efficiency vs frame for 20110608



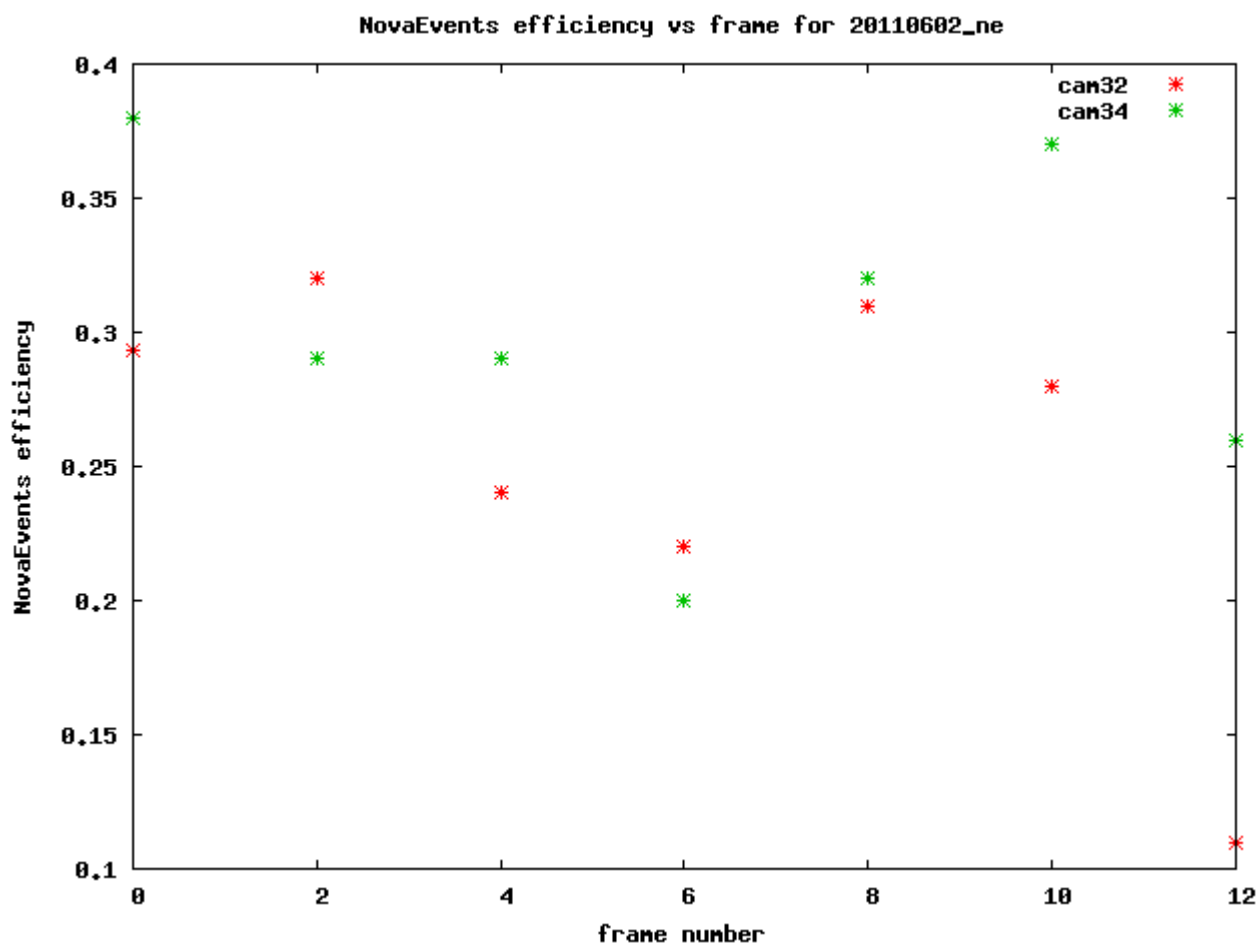
2. Efektywność flagi new_star w funkcji klatki dla wybranych kamer.



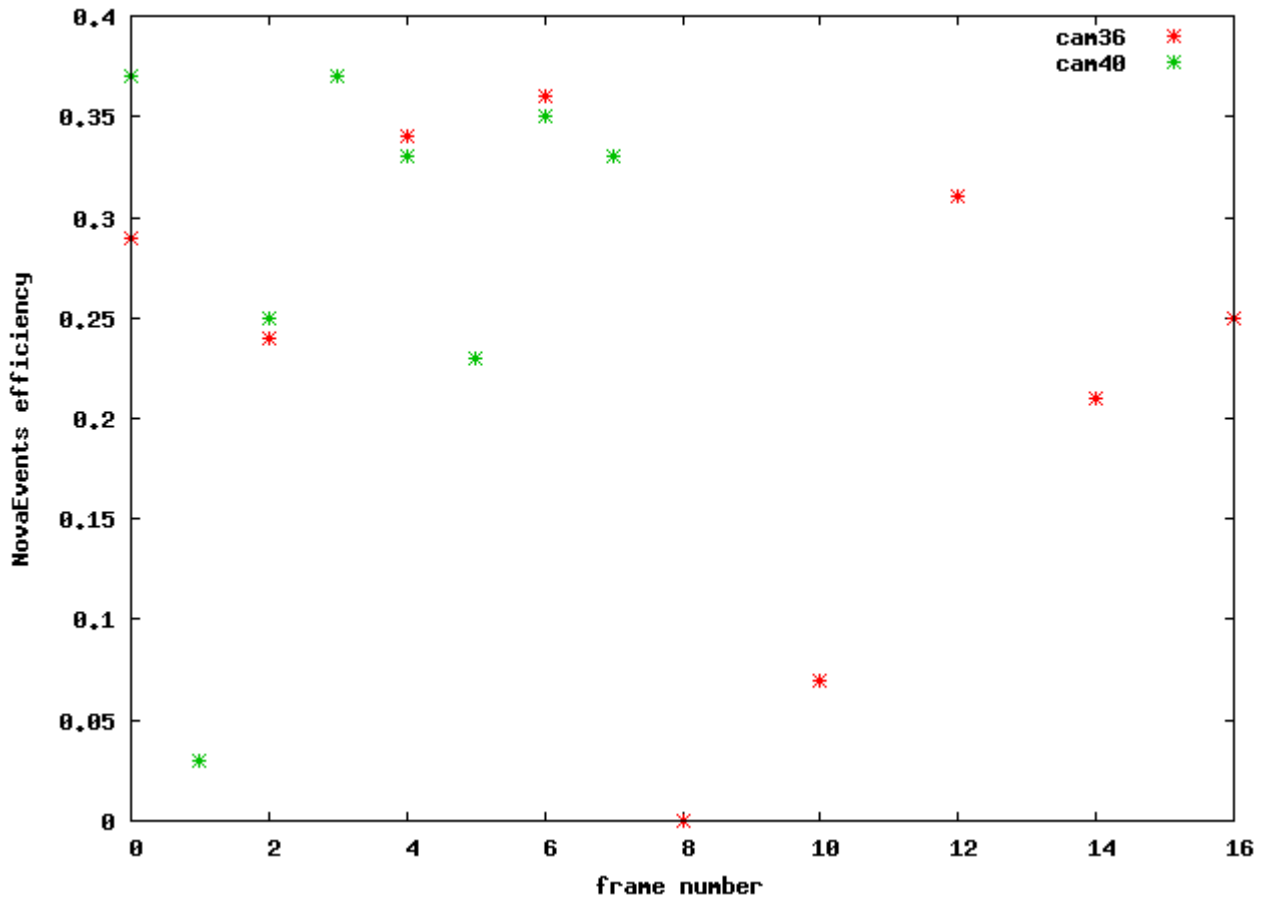
3. Efektywność wykrywania nowych w funkcji klatki dla min_obs_field=10



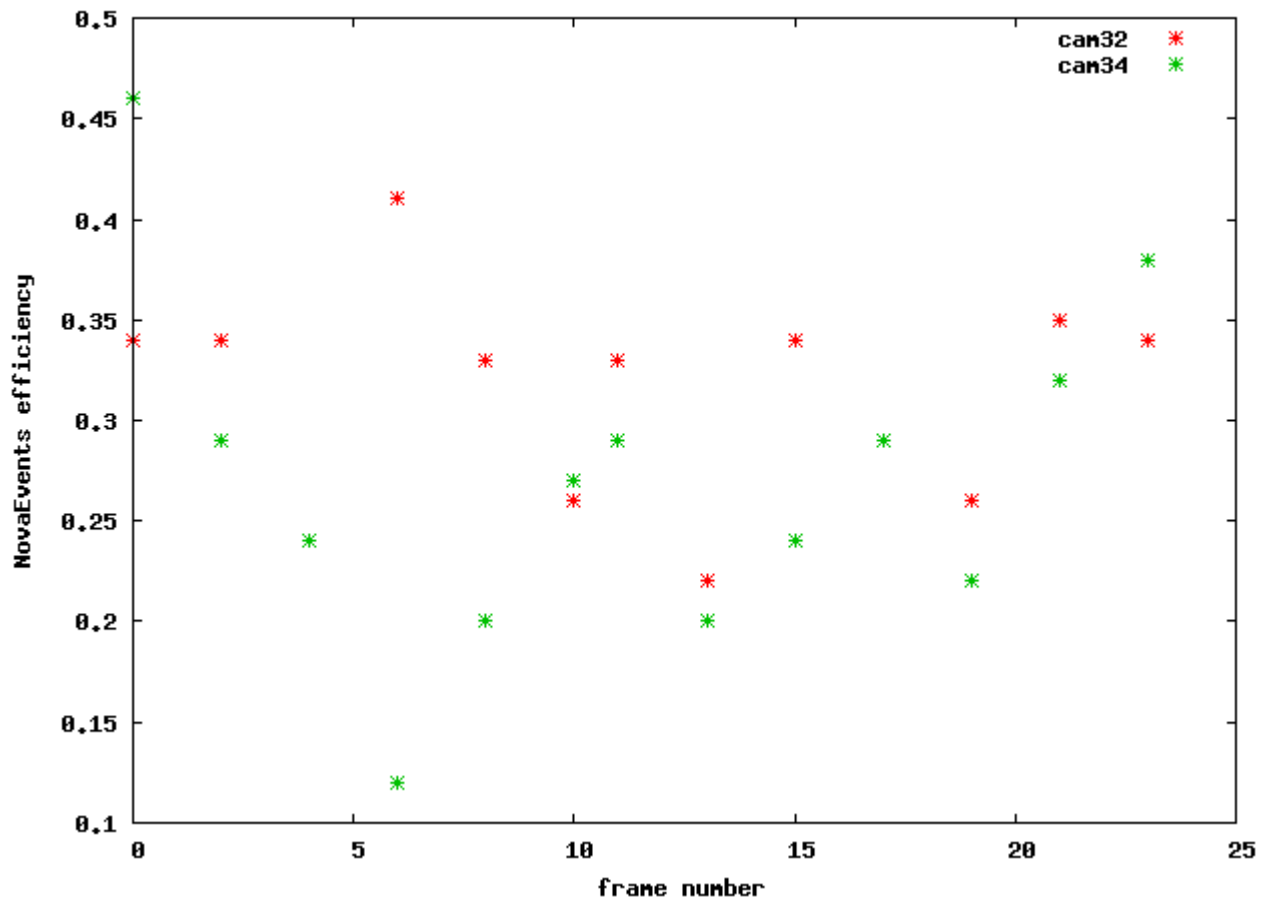
4. Efektywność wykrywania nowych w funkcji klatki dla min_obs_field=0



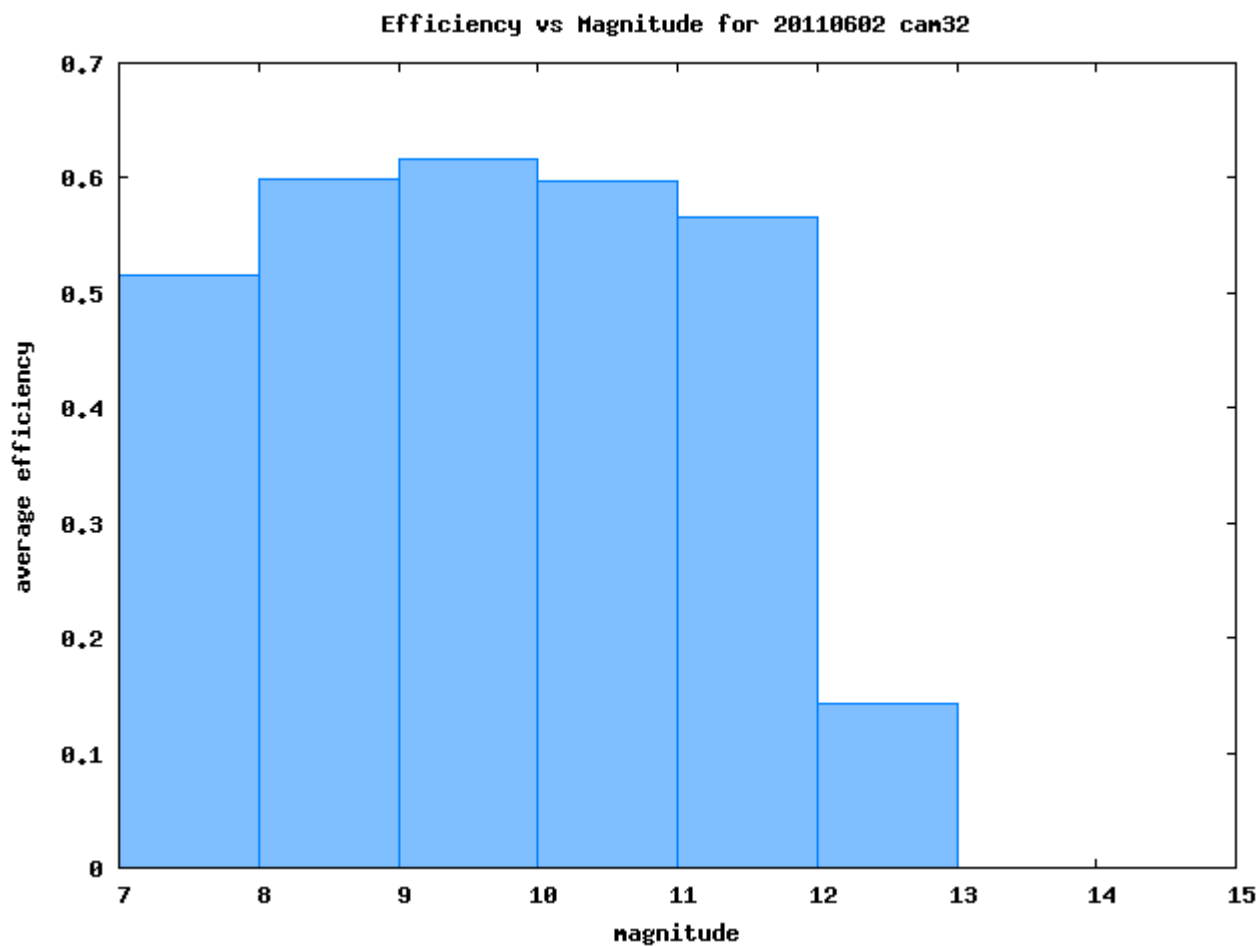
NovaEvents efficiency vs frame for 20110603_ne



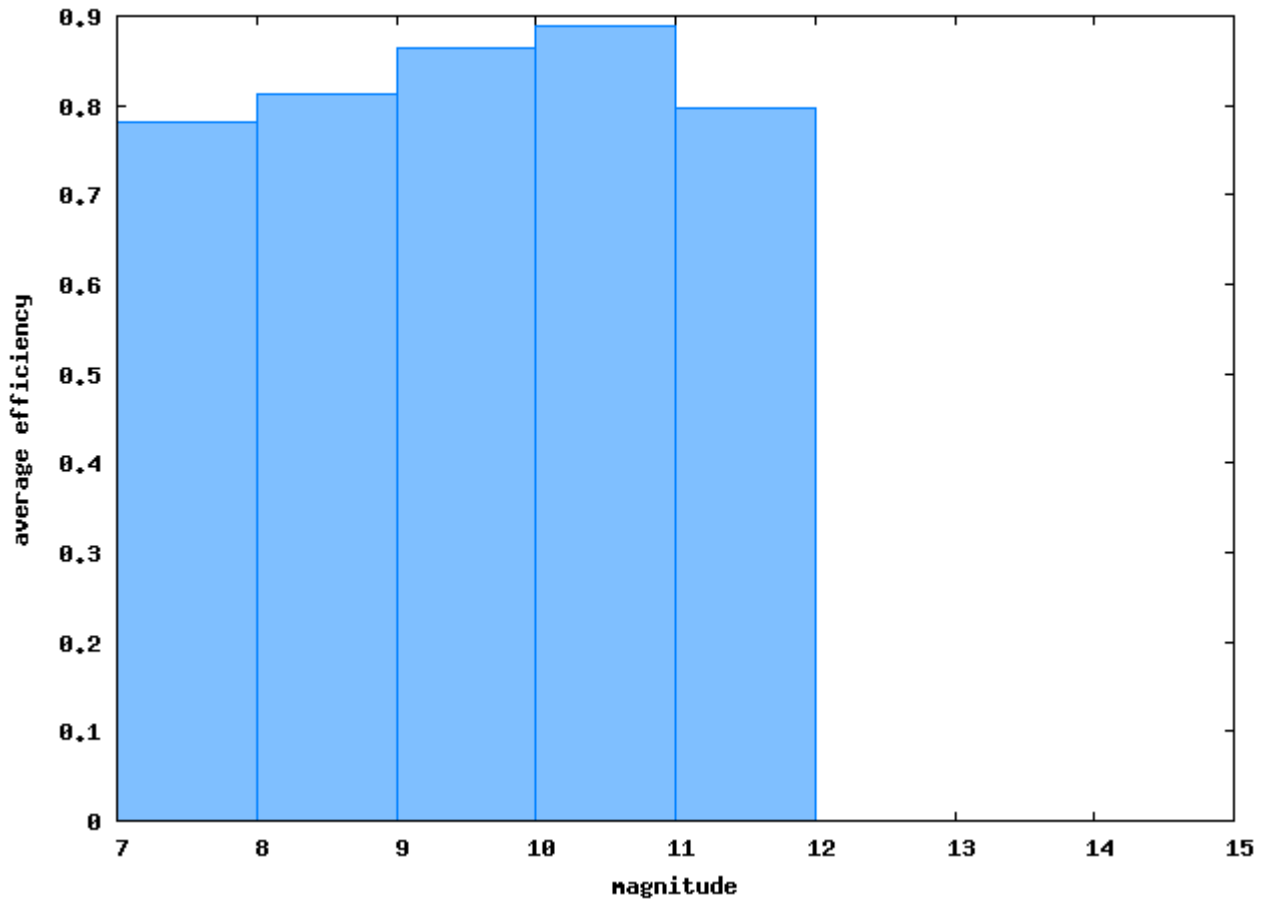
NovaEvents efficiency vs frame for 20110604_ne



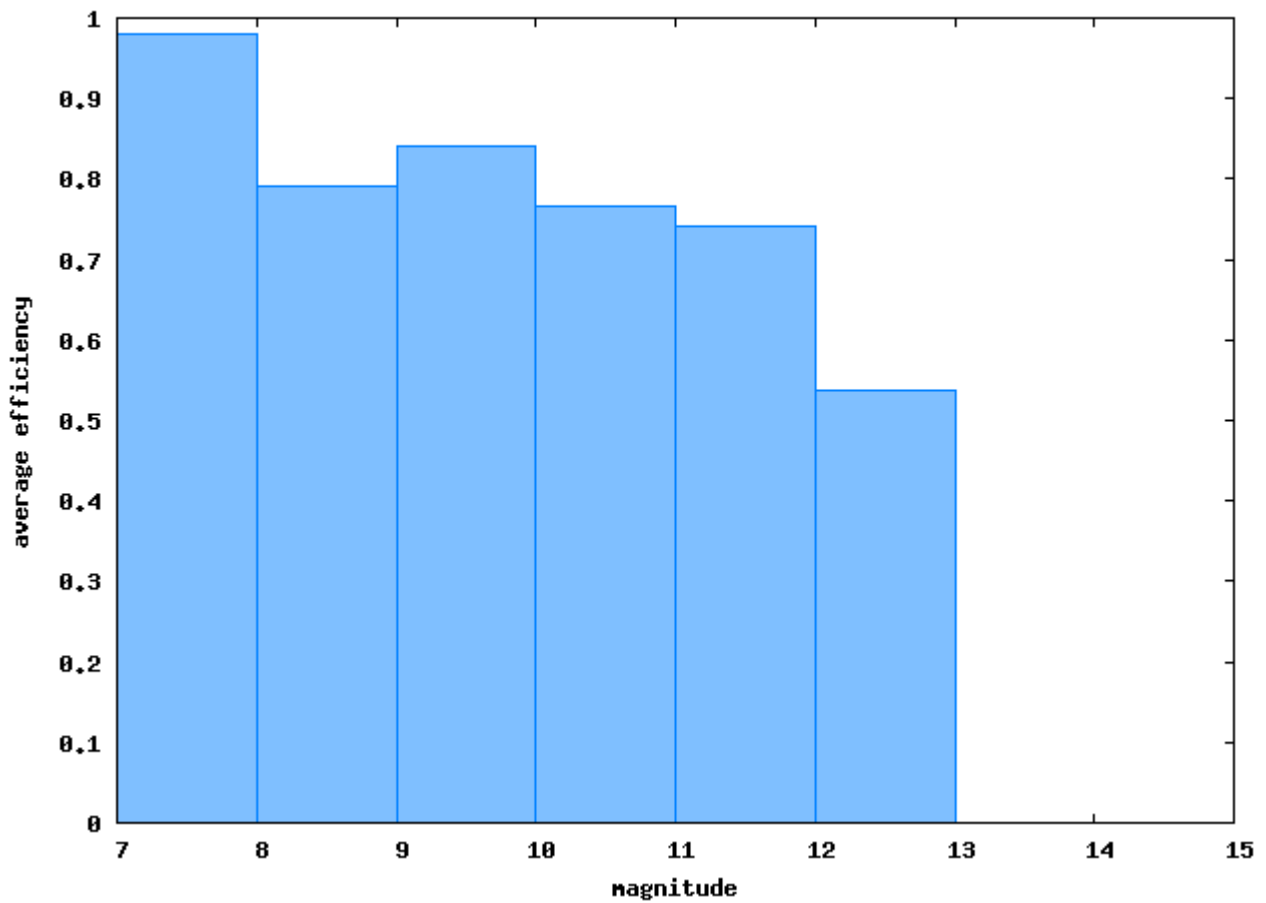
5. Efektywność astrometrii oraz wykrywania nowych (flaga new_star) w funkcji jasności.



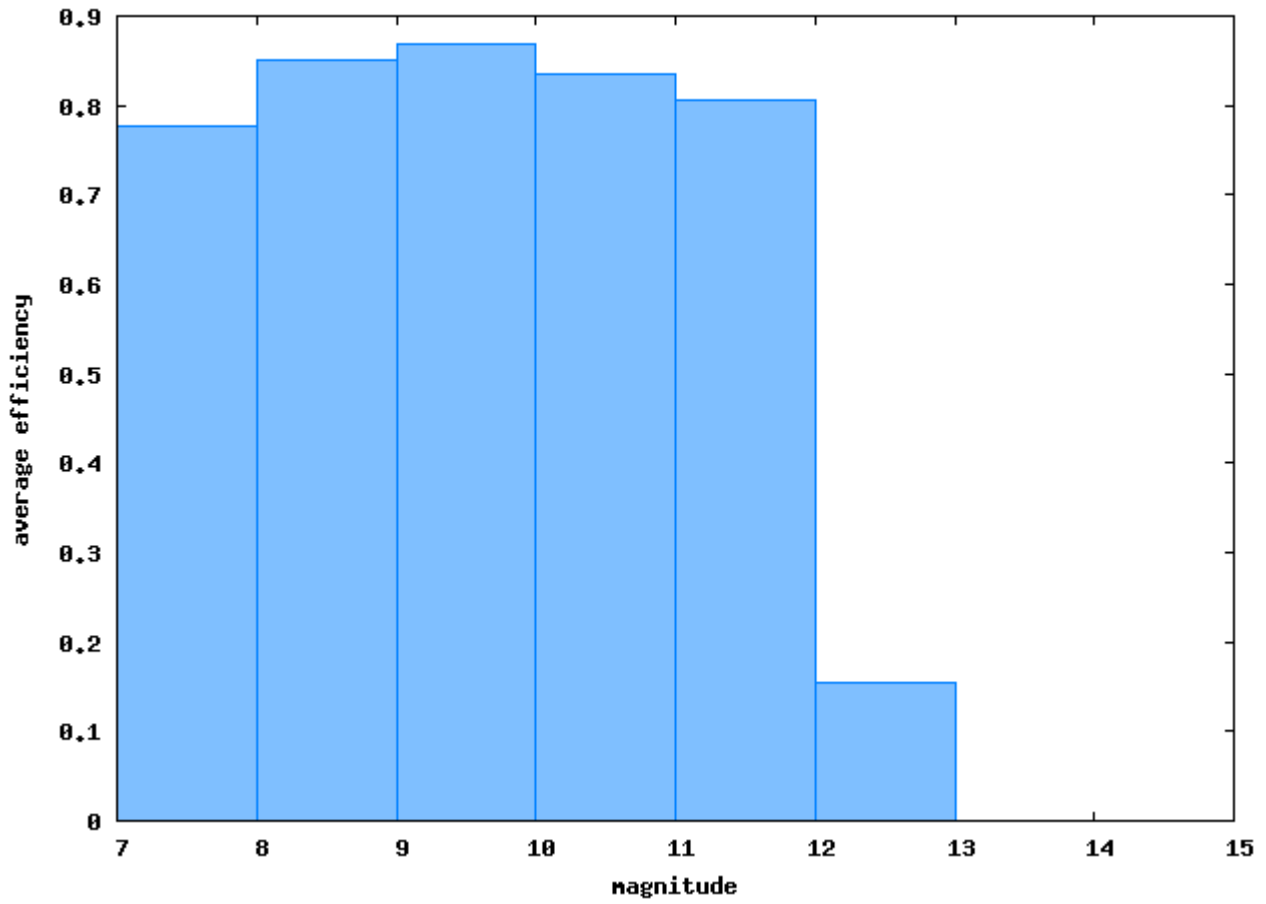
AstFile Efficiency vs Magnitude for 20110602 can40



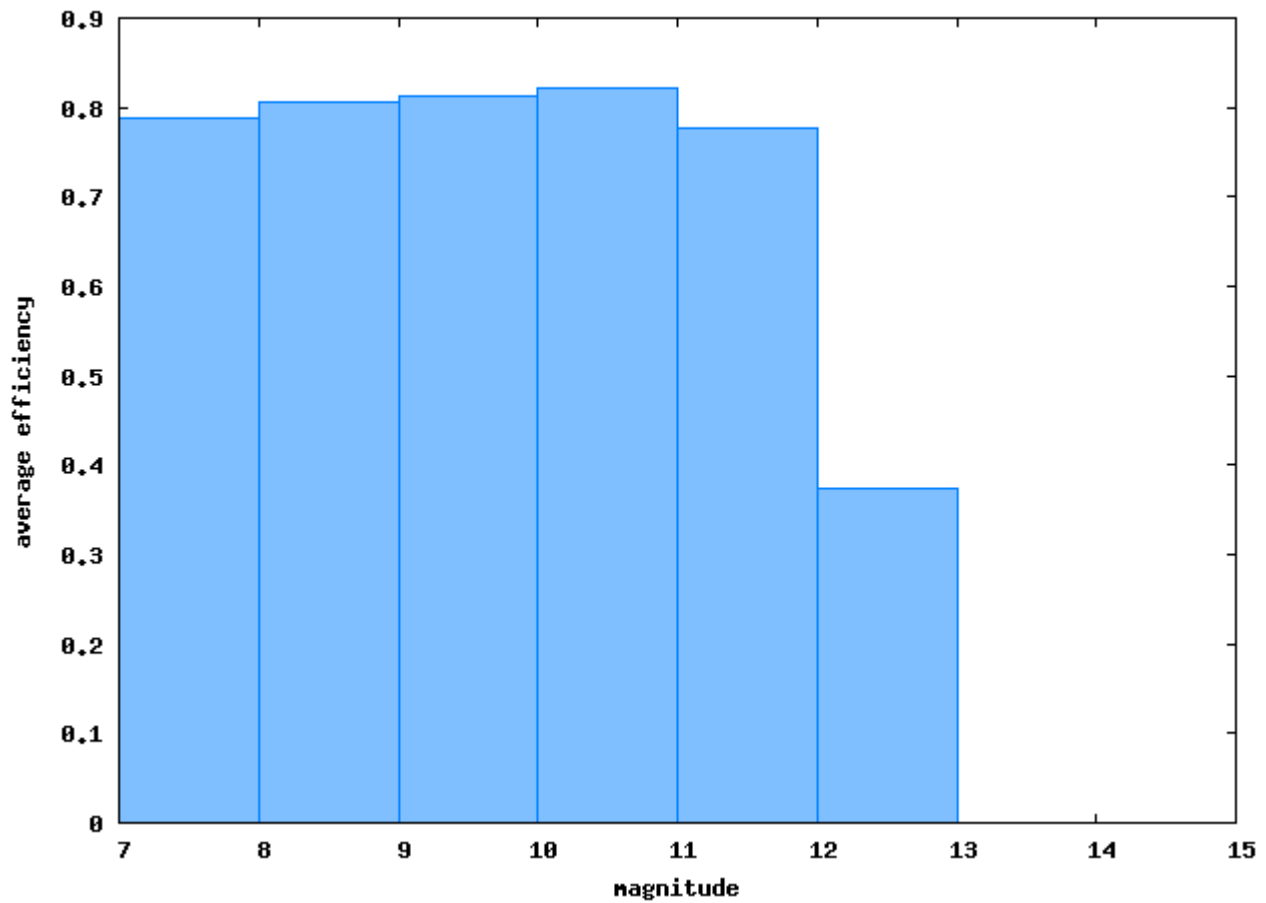
AstFile Efficiency vs Magnitude for 20110608 can36



AstFile Efficiency vs Magnitude for 20110604 can36



AstFile Efficiency vs Magnitude for 20110603 can40



7. Podsumowanie

Otrzymane wyniki pokazują, że straty w wydajności szukania gwiazd nowych związane z astrometrią wahają się w granicach 10% – 40%. Nie jest widoczna zależność od jasności w zakresie od 7 do 12 mag; powyżej tej granicy widoczny jest znaczny spadek wydajności w zakresie 12-13, obiekty ciemniejsze od 13 mag nie są zarejestrowane.

Wydajność flagowania obiektów jako nowych wykazuje podobną zależność od jasności, co w przypadku astrometrii. Odchylenia od średniej wydajności – co widać na wykresach – są jednak znacznie większe.

W otrzymanych danych wykrywanie nowych odbywa się z wydajnością wahającą się od 0 do ok. 40-50%, co jest znacznym spadkiem w stosunku do strat wynikających z działania procedury astrometrii. Wydajność równa zero występuje na kilku klatkach w przypadku, gdy algorytm wywoływany jest z parametrem `min_obs_field=10`. Analiza logów wynikowych algorytmu pozwala odnaleźć przyczynę – przykładowy wypis:

```
Star star=8721062 skipped, number of field observation too low : 3 < 10 (
id_frm=7412 )
```

Jest nią zbyt mała liczba wykonanych dla danego obszaru obserwacji; można to wyeliminować zmniejszając wartość parametru `min_obs_field`.

Dla `min_obs_field=0` również widoczna jest zerowa efektywność dla `cam36_110603_008`. W logu, w miejscu dotyczącym tej klatki wypisane są komunikaty typu:

```
star=9036418 - CUT 1 OK : number of field obs 12 >= 0 ( id_frm=7483 )
star=9036418 - CUT 2 FAILED : rejected due to CLOUDS_CHECK avg=2303 >
2000
LOG : star=9036418 : rej_tycho = 7
(...)
Event rejected ( rej_tycho = 7 ) and not saved to DB ( star = 9036418 )
```

Co oznacza odrzucenie ze względu na zachmurzenie nieba.

8. Bibliografia

[1] Marcin Sokołowski, *Investigation of astrophysical phenomena in short time scales with „Pi of the Sky” apparatus*, Warsaw 2007.