Zè's Google Summer of Code Application

Zè Vinícius (`@mirca`)
`http://mirca.github.io`

March 25, 2016

1 Sub-organization

Astropy

2 Student Information

- Name: Zé Vinícius
- Email: jvmirca@gmail.com
- Telephone: +55 83 99654-8106
- Time zone: UTC $-03:00$
- Github name: `@mirca`
- Blog: `http://mirca.github.io`

3 University Information

- University: Federal University of Campina Grande, Brazil
- Major: Electrical Engineering
- Current Year and Expected Graduation data: 4th year, 2017
- Degree: BSc

4 Student Background

I am a senior year undergraduate student in Electrical Engineering with plans to apply to a PhD program in Optics. I have received the Young Author Recognition Award by the International Telecommunications Union (ITU) twice.

During the Summer of 2015, I worked as an undergraduate guest researcher with the Nanofabrication Research Group (NRG) at the National Institute of Standards and Technology (NIST), USA. In that opportunity, I developed MATLAB code for single molecule localization microscopy (beyond diffraction limit).

Basically, I wrote a toolbox which provides to users easy access to maximum likelihood estimators (MLE) for the position, number of photon counts, width, and number of photons from background of dye molecules, and the uncertainties associated with each of the mentioned parameters as well. In other words, the code fits the point spread functions, which was assumed to have a symmetric Gaussian shape, to, in general, Poisson data (source plus background).

In addition, I wrote fitting routines which takes into account specific instrument information, e.g., electron multiplying CCD cameras (EMCCD) and scientific CMOS cameras. This was important in order to fully characterize the counts' statistics, so that the fitting procedure would give meaningful results even in low counts scenarios. For example, counts from EMCCD images are not exactly Poisson distributed.

Figure 1 shows an use case of the developed code. The dye molecules are localized in a EMCCD image using maximum likelihood point spread function fitting.

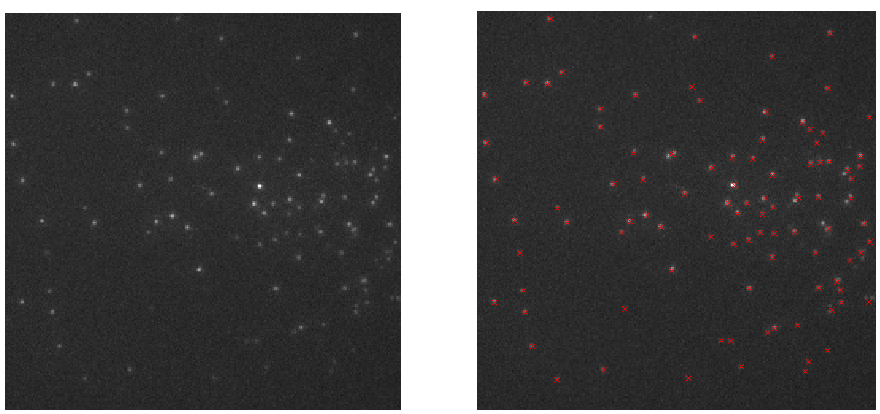

Figure 1: Left: EMCCD image of dye molecules. Right: The red markers represent the center position after performing PSF fitting with MLE. Uncertainty on the position is about $20nm$. The pixel width is $120nm$ and the EMCCD gain is 22.

Having said that, the problem of single molecule localization is, surprisingly, quite similar with the project that I am applying for, namely, "Implement PSF photometry for fitting several overlapping objects at once", with Brigitta Sipocz and Moritz Güenther as mentors.

I am quite sure the Google Summer of Code will be a very nice opportunity not only to become more involved and work closer with `photuilts` (`astropy`), but also to learn a lot and improve my coding skills, and to write readable, maintainable, and high quality code.

4.1 Relevant Skills, Awards, and Experience

Research oriented and problem solver student I have 3+ years of experience as undergraduate research assistant (I have published about ten articles in international conferences and three papers in journals). Currently, I hold an undergraduate teaching assistant scholarship for the course of Probability and Statistics for Electrical Engineering and Computer Science.

Programming and Open Source Development Roughly 1+ year of experience with Python, 3+ years of experience with C/C++, and 3+ months with Github.

Coursework During an one year period, I was a visiting student at the Catholic Univeristy of America and the University of Maryland, College Park, USA, in which I took graduate level coursework in Data Analysis, Stochastic Processes, and Estimation and Detection Theory.

Awards

- Travel Grant – IEEE Antennas and Propagation Society, 2016
- Probability and Statistics Teaching Assistantship – Spring 2016
- Undergraduate Guest Researcher – Summer 2015, NIST, USA
- Fully funded scholarship recipient – The Catholic University of America, USA, Fall 2014 (GPA: 4.0) and Spring 2015 (CGPA: 3.54)
- Young Author Recognition – International Telecommunication Union, 2014 and 2015
- Undergraduate Research Assistantship – Institute for Advanced Studies in Communications, Brazil, 2012 – 2014

5 Project Title

Implement PSF photometry for fitting several overlapping objects at once

6 Proposal Title

PSF photometry for fitting multiple overlapping objects simultaneously

7 Proposal Description

7.1 Project Abstract

Aperture photometry assumes that the background varies in a linear fashion in the aperture's vicinity. However, in a dense star cluster the background is usually nonlinear. Therefore, one may use point spread function (PSF) photometry in order to meaningfully measure the brightnesses of the sources. In the latter approach, a single model (usually Gaussian shaped) is fitted to each object allowing one to determine, with subpixel precision, their position (x, y center), amplitude, and width. The uncertainties associated with each of these parameters may be determined as well.

However, this becomes a non straightforward task when considering fitting multiple overlapping objects. To do so, one can not "just fit a model with hundreds parameters". In fact, there exist several problems with this "brute force" approach, and the most critical one might be that the parameter space will have many dimensions (as many as the number of parameters, precisely), which almost certainly will make optimization algorithms to diverge or to get stuck in a local minima.

Hence, my primary task is to work on developing an algorithm to localize, fit, and perform photometry of several overlapping objects (e.g. a dense star cluster, globular clusters, etc) simultaneously.

8 Proposal Timeline

For clarification, the community bonding period starts on April 22, the first week of coding starts on May 23, and the last week of coding starts on August 23. I'm not committed with any other activity during summer, however I will be attending a conference in the United States from June 26 to July 1. Writing tests and documentation are implicitly included on the proposed timeline.

- Community Bonding Period
  - Meeting with mentors via Skype. Meeting other photutils and astropy core developers, and the develop community in general, in order to gather suggestions to build a concrete plan of action before actually starting writing code.
  - Perform a thorough bibliographical research to see how astronomy softwares have been tackling the problem of fitting multiple overlapping sources. For now, the classic DAOPHOT [1] fitting routines for crowded fields will certainly be implemented.
  - Discuss design API and interface.
- Week 1
  - Get acquainted with PSF photometry features of photutils.
- Weeks 2-4
  - Implement and validate a similar algorithm as presented in the classic DAOPHOT fitting routines for crowded fields.
- Week 5
  - Prepare for midterm evaluation. Start to implement and validate at least two recent published PSF photometry algorithms for crowded fields.
- Week 6
  - I will be attending a conference this week. I will work on Sundays to make up for this.
- Weeks 7-8
  - Implement and validate at least two recent published PSF photometry algorithms for crowded fields (continuation).
- Week 9
  - Implement benchmarking and PSF uncertainties functions.
  - Starting cythonize or parallelize any code, if necessary.
- Week 10

- Compare the performance of the implemented PSF photometry algorithms with respect to computational time, uncertainties, number of correctly detected sources, etc, using artificial and real data (e.g., Spitzer).
- Continue to cythonize or parallelize any code, if necessary.

- Weeks 11-12
  - Write notebooks with examples/tutorials demonstrating the use of the developed fitting procedures.
- Weeks 13-14
  - Improve documentation, improve tests, debug, and final evaluation.
- If time allows
  - Implement camera-specific (CCD, EMCCD, and sCMOS) maximum likelihood fitters, including uncertainties from the Fisher Information Matrix. See [2–4]. This is particularly interesting in extreme scenarios where the photon counts are too low and the underlying statistics of the observations are not Gaussian distributed.
  - Research and implement other source detection algorithms, PSF models, PRF photometry, optimization algorithms, etc...
  - Defocused PSF photometry
- *A Posteriori* GSoC
  - Keep sending PRs to Astropy.

9 Link to patches

- Implemented Jackknife resampling method:
  `https://github.com/astropy/astropy/pull/4439`
- Implemented circular statistics with support to Quantity:
  `https://github.com/astropy/astropy/pull/4472`
- Fixed minor issues in astropy.units doc page:
  `https://github.com/astropy/astropy/pull/4527`
- Implementing Akaike and Bayesian Information Criterions:
  `https://github.com/astropy/astropy/pull/4716`

References

[1] P. B. Stetson. DAOPHOT - A computer program for crowded-field stellar photometry. *Publications of the Astronomical Society of the Pacific*, 99:191–222, March 1987.

[2] J. Skottfelt, D. M. Bramich, R. Figuera Jaimes, U. G. Jørgensen, N. Kains, K. B. W. Harpsøe, C. Liebig, M. T. Penny, K. A. Alsubai, J. M. Andersen, V. Bozza, P. Browne, S. Calchi Novati, Y. Damerdji, C. Diehl, M. Dominik, A. Elyiv, E. Giannini, F. Hessman, T. C. Hinse, M. Hundertmark, D. Juncher, E. Kerins, H. Korhonen, L. Mancini, R. Martin, M. Rabus, S. Rahvar,

G. Scarpetta, J. Southworth, C. Snodgrass, R. A. Street, J. Surdej, J. Tregloan-Reed, C. Vilela, and A. Williams. EMCCD photometry reveals two new variable stars in the crowded central region of the globular cluster NGC 6981. *Astronomy & Astrophysics*, 553:A111, May 2013.

[3] Roger Smith. Astronomical demands on cmos imaging sensors. The Workshop on CMOS Applications in Astronomy and Space Sciences, 2011.

[4] P. Qiu, Y.-N. Mao, X.-M. Lu, E. Xiang, and X.-J. Jiang. Evaluation of a scientific CMOS camera for astronomical observations. *Research in Astronomy and Astrophysics*, 13:615–628, May 2013.