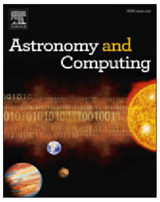
Astronomy and Computing 24 (2018) 129–142

Contents lists available at ScienceDirect

Astronomy and Computing

journal homepage: www.elsevier.com/locate/ascom



Full length article

SCARLET: Source separation in multi-band images by Constrained Matrix Factorization[☆]

P. Melchior[a,*], F. Moolekamp[a], M. Jerdee[a], R. Armstrong[a], A.-L. Sun[b], J. Bosch[a],
R. Lupton[a]

[a] Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA
[b] Department of Physics & Astronomy, Johns Hopkins University, Bloomberg Center, 3400 N. Charles St., Baltimore, MD 21218, USA

ARTICLE INFO

Article history:
Received 7 March 2018
Accepted 10 July 2018
Available online 19 July 2018

Keywords:
Methods
Data analysis
Techniques
Image processing
Galaxies
Non-negative matrix factorization

ABSTRACT

We present the source separation framework SCARLET for multi-band images, which is based on a generalization of the Non-negative Matrix Factorization to alternative and several simultaneous constraints. Our approach describes the observed scene as a mixture of components with compact spatial support and uniform spectra over their support. We present the algorithm to perform the matrix factorization and introduce constraints that are useful for optical images of stars and distinct stellar populations in galaxies, in particular symmetry and monotonicity with respect to the source peak position. We also derive the treatment of correlated noise and convolutions with band-dependent point spread functions, rendering our approach applicable to coadded images observed under variable seeing conditions. SCARLET thus yields a PSF-matched photometry measurement with an optimally chosen weight function given by the mean morphology in all available bands. We demonstrate the performance of SCARLET for deblending crowded extragalactic scenes and on an AGN jet–host galaxy separation problem in deep 5-band imaging from the Hyper Suprime-Cam Strategic Survey Program. Using simulations with prominent crowding we show that SCARLET yields superior results to the HSC-SDSS deblender for the recovery of total fluxes, colors, and morphologies. Due to its non-parametric nature, a conceptual limitation of SCARLET is its sensitivity to undetected sources or multiple stellar population within detected sources, but an iterative strategy that adds components at the location of significant residuals appears promising. The code is implemented in Python with C++ extensions and is available at https://github.com/fred3m/scarlet.

1. Introduction

Modern astronomical wide-field surveys cover large areas of the sky at ever increasing depths, revealing more objects and low-surface-brightness features of extended objects that were previously too faint to detect. These gains drive investigations into galactic, extragalactic and cosmological phenomena at an unprecedented level of detail and statistical power. On the other hand, because of the enhanced sensitivity, a larger fraction of the observed area is associated with detectable objects, thereby increasing the chance that multiple objects overlap. This so-called "blending" constitutes a major concern for the analysis of existing and upcoming surveys, especially those that observe from the ground.

The majority of methods for measuring the properties of celestial objects assume that every object can be considered isolated. If that assumption holds, well-defined and accurate measurements of the flux, position, shape, and morphology can routinely be made with methods that are either based on moments of the light distribution within some aperture or on parametric fits to the images.

However, the notion of isolated objects is becoming increasingly obsolete. With a limiting magnitude of $i \approx 24$, DES[1] (Dark Energy Survey Collaboration, 2016) finds that 30% of galaxies suitable for weak-lensing measurements are affected by blending (Samuroff et al., 2018). For HSC[2] (Aihara et al., 2018a), whose Wide survey has a limiting magnitude of $i \approx 26$, Bosch et al. (2018) find that 58% of measured objects are in blended groups, a dramatic increase despite a substantially better average seeing than DES. LSST[3] (Ivezic et al., 2008) expects to reach $i \approx 27$ after 10 years of operations , and it is estimated that 63% of observed galaxies will have Sérsic model photometry that is altered by more than 2% due to the presence of neighbors (Sanchez et al., in preparation). Even

---

[☆] This code is registered at the ASCL with the code entry ascl:1803.003.
[*] Corresponding author.

E-mail addresses: peter.melchior@princeton.edu (P. Melchior), fredem@princeton.edu (F. Moolekamp).

[1] https://www.darkenergysurvey.org.
[2] http://hsc.mtk.nao.ac.jp/ssp/.
[3] http://lsst.org.

https://doi.org/10.1016/j.ascom.2018.07.001
2213-1337/© 2018 Elsevier B.V. All rights reserved.

more problematically, in a comparison study of HST and Subaru imaging of a galaxy cluster field, where the Subaru data had similar seeing and depth so as to serve as a proxy for LSST, Dawson et al. (2016) found that 14% of observed galaxies are blended but not recognized as such in the ground-based images. Slightly larger numbers for unrecognized blends are found for HSC in a study that inserted fake objects into real survey images to infer how many of them could be recovered (Murata et al., in preparation).

This lack of separability between objects necessitates the employment of techniques that analyze entire scenes with overlapping objects. For direct measurements, such as moments within apertures, there is no accurate way to correct for the excess light from overlapping objects because such a correction depends e.g. on radial profiles of the objects involved, which cannot be determined well for blends. As a consequence, multiple objects need to be *modeled* either iteratively (by masking all but one) or simultaneously, often requiring sophisticated and fine-tuned schemes to prevent unstable or physically implausible solutions (e.g. Barden et al., 2012; Drlica-Wagner et al., 2018). Any such scheme is suitable to extract and separate the objects in celestial scenes, i.e. to "deblend" those scenes. Differences between the schemes include the propagation of errors, which conceptually favors simultaneous approaches, and whether the desired measurements are generated directly from deblender models or by passing them on to established measurement algorithms for isolated objects.

Traditional deblending approaches in astronomy are achromatic, i.e. they employ information from only a single image. SExtractor detects blending by thresholding an image at a range of intensity levels and searching for sets of pixels that are connected at a lower threshold but split into several connected regions at a higher one (Bertin and Arnouts, 1996). As a consequence of the splitting approach, the association of pixels to objects is unique and exclusive, i.e. in the internal representation of blended objects they do not overlap. This unrealistic notion has necessitated mitigation strategies or fine-tuning to prevent "over-deblending" of larger galaxies caused by smaller and fainter companions or interlopers (e.g. Rix et al., 2004).

The deblender in the SDSS *Photo* pipeline (Lupton, in preparation) does allow for overlap between nearby objects and, consequently, needs to estimate the portion of any pixel's flux that is due to each object. It uses a two-step approach, in which first a template is constructed for each object based on the requirement that pixel values symmetrically across the object's peak pixel be identical; they are generally not, so the minimum pixel value of those two pixels is adopted for both. Then, the original image values are projected onto those templates, associating each pixel's flux to different objects in proportion to the amplitude of the respective templates. Despite very few assumptions, the method mostly separates sources into physically plausible objects but struggles with situations where a central object is symmetrically surrounded by neighbors, for instance a blend with three peaks in a row (Lupton, in preparation).

More recently, Zhang et al. (2015) and Connor et al. (2017) proposed variants of inpainting techniques, where the relevant pixels of blended objects are replaced by an estimate of a local variable background, working inwards from an initially defined outline. The portion of the pixel flux above the background estimate is attributed to the respective object. These approaches implicitly account for blending by assuming that the background captures the flux contributions from neighboring objects. This is particularly useful when recovering small objects in multi-scale blending situations like the cores of galaxy clusters and removing them from the scene so that large objects can be measured separately.

While effective in many cases, all of the deblending schemes outlined above employ heuristic arguments for how to separate overlapping sources. They also perform the pixel–object association sequentially, one object at a time, thus losing the advantages of a simultaneous solution, for instance the ability to explore the degeneracies that arise because the objects are not isolated. However, it is our opinion that the biggest limitation stems from the restriction to a single image, and therefore a single filter band, while most modern surveys observe the sky in several filters. A visual inspection of multi-band images clearly suggests that color can serve as a powerful discriminator between different objects, even with severe overlap (see Fig. 1).

MuSCADeT (Joseph et al., 2016) addresses both limitations by building a joint model of multi-band image data. As their model is non-parametric, the number of degrees of freedom is large, which leads to many possible degeneracies in the solutions, so they demand that the spatial distribution of each source be sparse in the starlet (a form of isotropic wavelet) domain. The resulting solutions extract preferentially compact features down to the noise level, using a set of previously identified colors for each feature. For applications to wide-field multi-band data, we cannot generally assume to a priori identify the color of an object in the scene because there might not be a single pixel whose color is uncontaminated by other objects. We therefore seek the ability to update both spectral and morphological characteristics of the objects. While sharing noticeable similarities with MuSCADeT regarding the use of a non-parametric constrained morphological model, one can consider our approach an extension that also updates the source spectra as well as a generalization that allows an arbitrary number of constraints to be placed on each source.

The outline of the paper is as follows: We introduce our approach, dubbed scarlet, in Section 2, demonstrate its performance on real data and simulations in Section 3, and conclude in Section 4.

2. Methodology

We base our deblender on the assumption that astronomical scenes are superpositions of multiple components, each with

1. a spatially compact support and
2. a constant spectrum over that support.

For stellar fields this is obviously true, but even for galaxies, especially marginally resolved ones, which constitute the vast majority of galaxies in deep surveys, the assumption is appropriate. In addition, even large, extended galaxies can be thought of as conglomerations of components (e.g. bulges, discs, bars, star forming regions) for which the assumptions above hold at least approximately.[4] For instance, the popular bulge-disc decomposition for galaxies (e.g. MacArthur et al., 2003) is justified by this interpretation. We note that the assumption of linear superposition implies that components do not interact, which is correct only for transparent emitters. Absorption, e.g. by dust in the galaxy, can be approximated by allowing negative values in the source spectrum, but substantial opacities cannot fully be modeled because the effect depends on the amount of absorbing material *and* the intensity of the background radiation.

The assumptions above appear to lend themselves to a parametric modeling framework, where one assumes to know the shapes of the components and potentially their intrinsic spectrum, exploiting quite tight relations between colors and morphologies exhibited by galaxies in the late universe (e.g. Conselice, 1997; Ball et al., 2008). While drastically reducing the number of optimization parameters, we are critical of this approach for two reasons: First, in the translation of an intrinsic, restframe spectrum to the observed

---

[4] In the literature the terms "source separation" and "component separation" are often used interchangeably. To better reflect the hierarchical nature of astrophysical scenes, we will define the term "Source" as a collection of co-centered components that belong to the same astrophysical object.