# Image analysis and statistics: an introduction using *R* and RIPA

**Talita Perciano**[1,⋆]

1. Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States
⋆Contact author: tperciano@lbl.gov

**Keywords:** *R*, statistics, image analysis, **RIPA**, high performance

*R* is a public domain, open source, language and environment for statistical computing and graphics [4]. As a tool, *R* is very popular among scientists because it is free; it includes a diverse set of algorithms, and entails a large and broad community that participates in augmenting the tool. *R* provides a large variety of statistical techniques such as linear and non-linear models, classical statistical tests, time series models, classification and clustering. Moreover, it offers various graphical techniques and it is highly extensible. *R* is an integrated collection of software packages for data manipulation, performing calculations and graphics production.

The field of digital image processing refers to any process applied to a digital image performed by a computer [2]. A 2D image can be defined as a bidimensional function $f(x, y)$, where $x$ and $y$ are the spatial coordinates, and the amplitude of $f$ at a coordinate $(x, y)$ is called the intensity or the gray level of the image at that point. An image has a finite number of elements, and each one of them has a particular position and a scalar or vectorial value. For instance, a gray level image has a unique value for each position. This definition can be easily extended to $n$-dimensional images. This field of study is very broad, comprising numerous operations, from common ones such as contrast enhancement and noise reduction, to more complex ones such as image segmentation/classification and pattern recognition. The range of applications which can take advantage of image processing techniques is immense and its importance to scientific research increases drastically. Moreover, the advent of Big Data and Data Science [3] demands the use and development of such techniques in order to discover knowledge from datasets in several research fields: Geoscience and Remote Sensing, Bioinformatics, Biology, Medicine, Physics, Astronomy, Geology, to name a few.

**RIPA** [5] is a user friendly package that provides several image processing tools, which can be applied to different types of images such as binary, gray level, color and multispectral images. Analysis and exploration tools, such as those provided by **RIPA** can be applied to diverse domains and can be very useful and important in many challenging imaging problems. Besides a general tool that can be used in such manner, **RIPA** is presented here also with the purpose to provide an auxiliary tool for learning statistics and image analysis. This package, along with the book entitled "Introduction to Image Processing Using *R*: learning by examples" [1], has been used as a text book around the world for disciplines such as "introduction to image processing" and "introduction to statistics". As its new version comes up with new image analysis tools and with high performance additions, it becomes a valuable tool both for teaching and scientific research purposes involving Big Data.

## References

[1] Frery, A. C. and T. Perciano (2013). *Introduction to Image Processing Using R: learning by examples (SpringerBriefs)* (1 ed.). Springer. ISBN: 978-1447149491.

[2] Gonzalez, R. C. and R. E. Woods (2008). *Digital Image Processing* (3 ed.). Prentice Hall.

[3] Hey, T., S. Tansley, and K. Tolle (Eds.) (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, Washington: Microsoft Research.

[4] Hornik, K. (2002). The R Project for Statistical Computing. http://www.r-project.org. Accessed March 25, 2014.

[5] Perciano, T. and A. C. Frery (2009). R Image Processing and Analysis. http://cran.r-project.org/src/contrib/Archive/ripa/. Accessed March 25, 2014.