

Using the spotSegmentation package

This document illustrates the use of the `spotSegmentation` package (Li et al. 2005) on the `spotSegTest` dataset (also available in the package). The `spotSegmentation` package consists of two basic functions:

`spotgrid`: determines spot locations in blocks within microarray slides

`spotseg`: determines foreground and background signals within individual spots

These functions will be illustrated on the `spotSegTest` dataset, which consists of a portion of the first block from the first microarray slide image from van't Wout et al. (2003).¹ This data set is a data frame, with two columns, one from each of the two channels of absorption intensities. The `spotSegTest` dataset can be obtained via the `data` command once the `spotSegmentation` package is installed.

```
> data(spotSegTest)
```

The data are encoded for compact storage, and they need to be transformed as follows in order to extract the intensities:

```
> dataTransformation <- function(x) (256*256-1-x)^2*4.71542407E-05
> chan1 <- matrix(dataTransformation(spotSegTest[,1]), 144, 199)
> chan2 <- matrix(dataTransformation(spotSegTest[,2]), 144, 199)
```

Note that this transformation is specific to this data; in general stored image data must be converted as needed to image intensities. Figures 1 and 2 show the image data for the two channels in reverse gray scale. The gray scale is calibrated using the logarithm of the intensities for better contrast. These plots can be obtained with the `spotSegmentation` package using the following commands:

```
> plotBlockImage(log(chan1))
> plotBlockImage(log(chan2))
```

¹The image data for the whole block is quite large in size, and is available as a separate R/Bioconductor package `HIVcDNAvantWout03`.

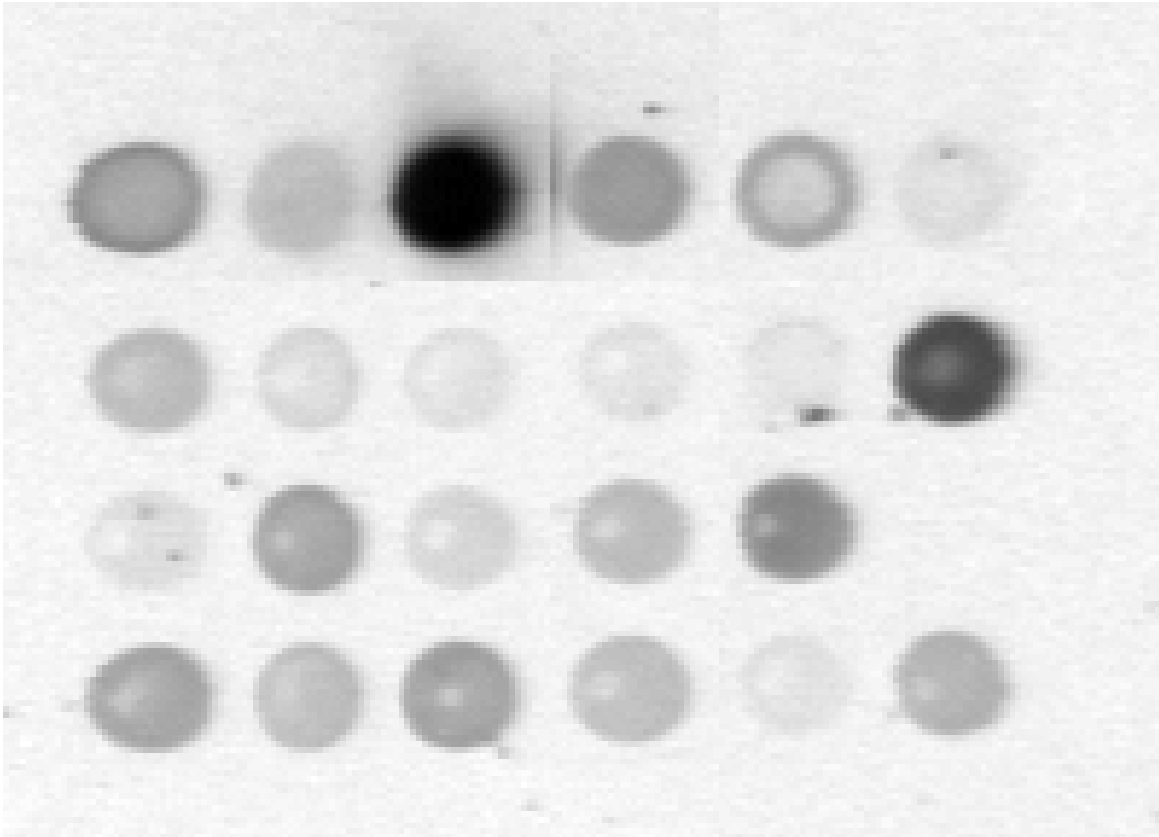


Figure 1: Reverse gray-scale plot of image intensities of a subset of spots from channel 1 (Cy3 green) of a portion of the first block from the first slide of HIV data from van't Wout et al. (2003).

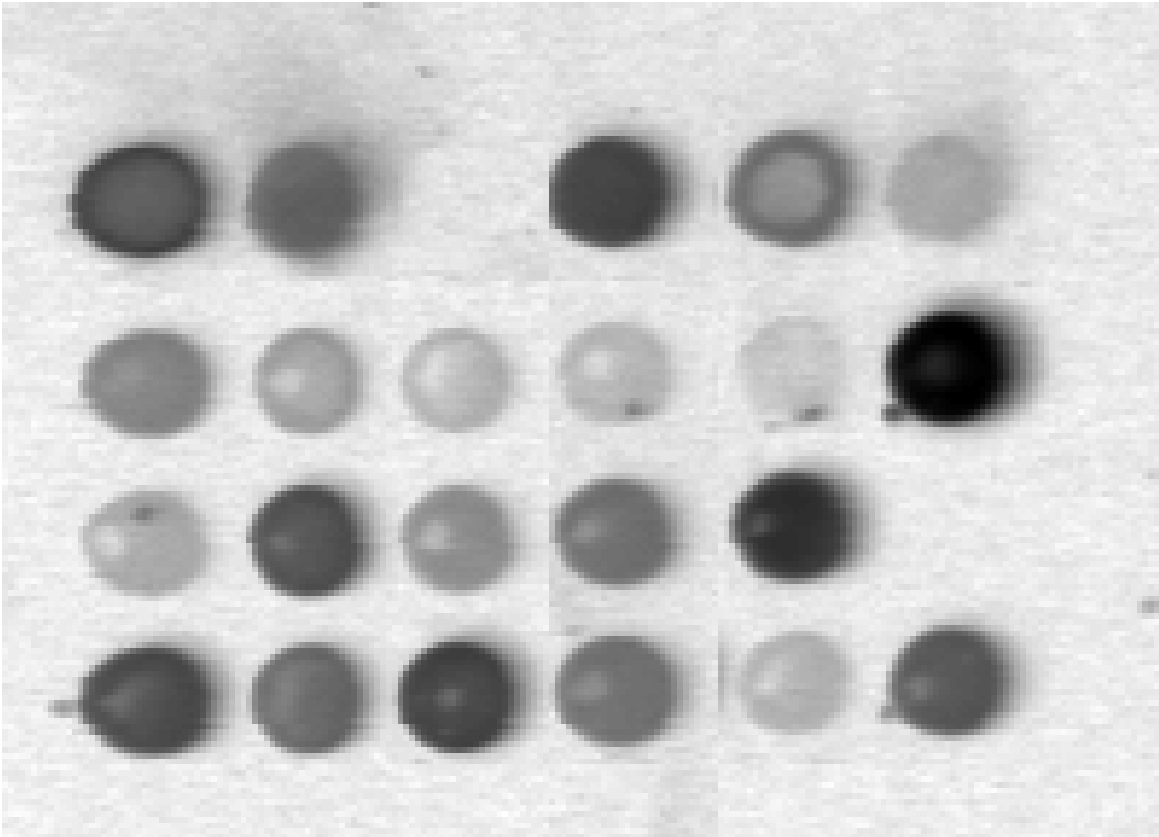


Figure 2: Reverse gray-scale plot of image intensities of a subset of spots from channel 2 (Cy5 red) of a portion of the first block from the first slide of HIV data from van't Wout et al. (2003).

The function `spotgrid` can be used to divide the microarray image block into a grid separating the individual spots.

```
> hivGrid <- spotgrid( chan1, chan2, rows = 4, cols = 5, show = TRUE)
```

```
> hivGrid
```

```
$rowcut
```

```
[1] 17 50 77 104 139
```

```
$colcut
```

```
[1] 12 41 66 94 123 151 183
```

Here we have used the knowlege that there are 4 rows and 6 columns in this subset of spots from the microarray image. The `show` option allows display of the image, shown in Figure 3.

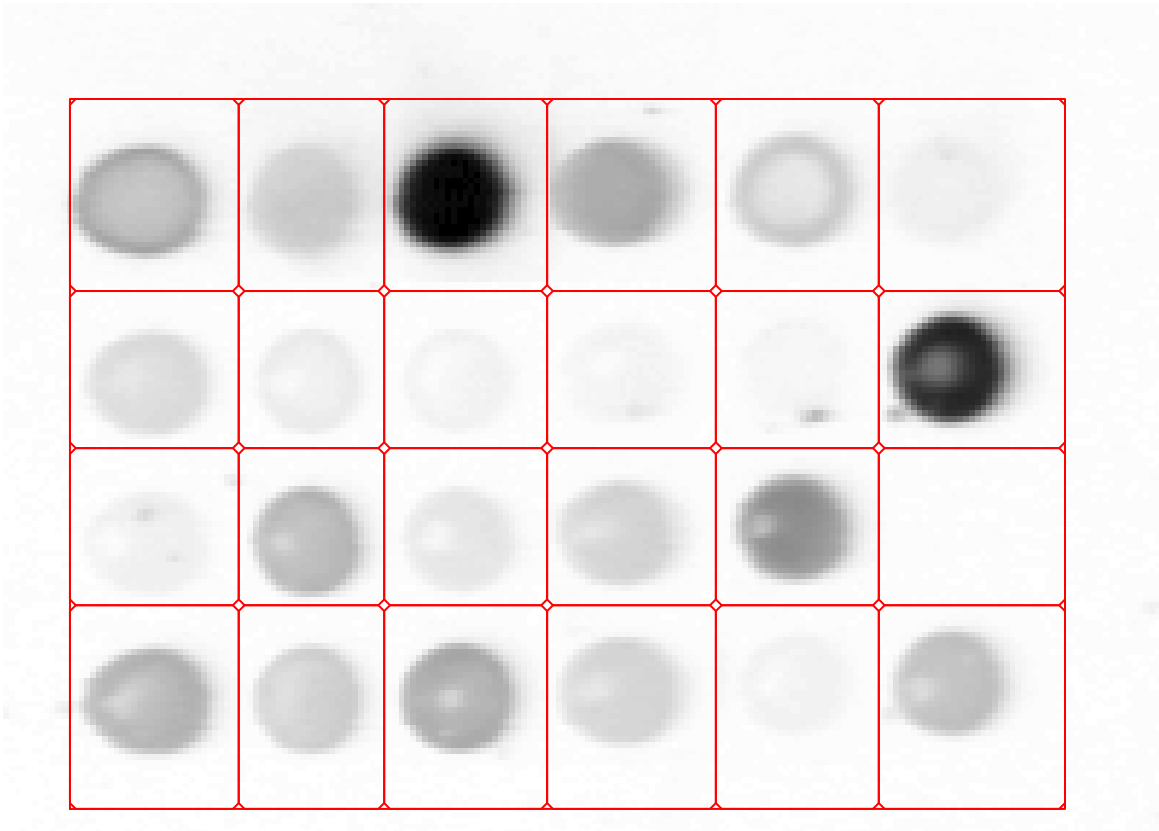


Figure 3: Grid delimiting microarray spots determined by function `spotgrid` for the HIV data subset. The underlying image is the sum of the intensities for the two channels.

The individual spots can now be segmented using the function `spotseg`. The following segments all spots in the block:

```
hivSeg <- spotseg( chan1, chan2, hivGrid$rowcut, hivGrid$colcut)
```

```
plot(hivSeg)
```

The corresponding plot is shown in Figure 4.

The segmentation is achieved using the `mclust` package for model-based clustering (Fraley and Raftery, 1999, 2003, 2006) , followed by a connected component analysis for removal of artifacts.

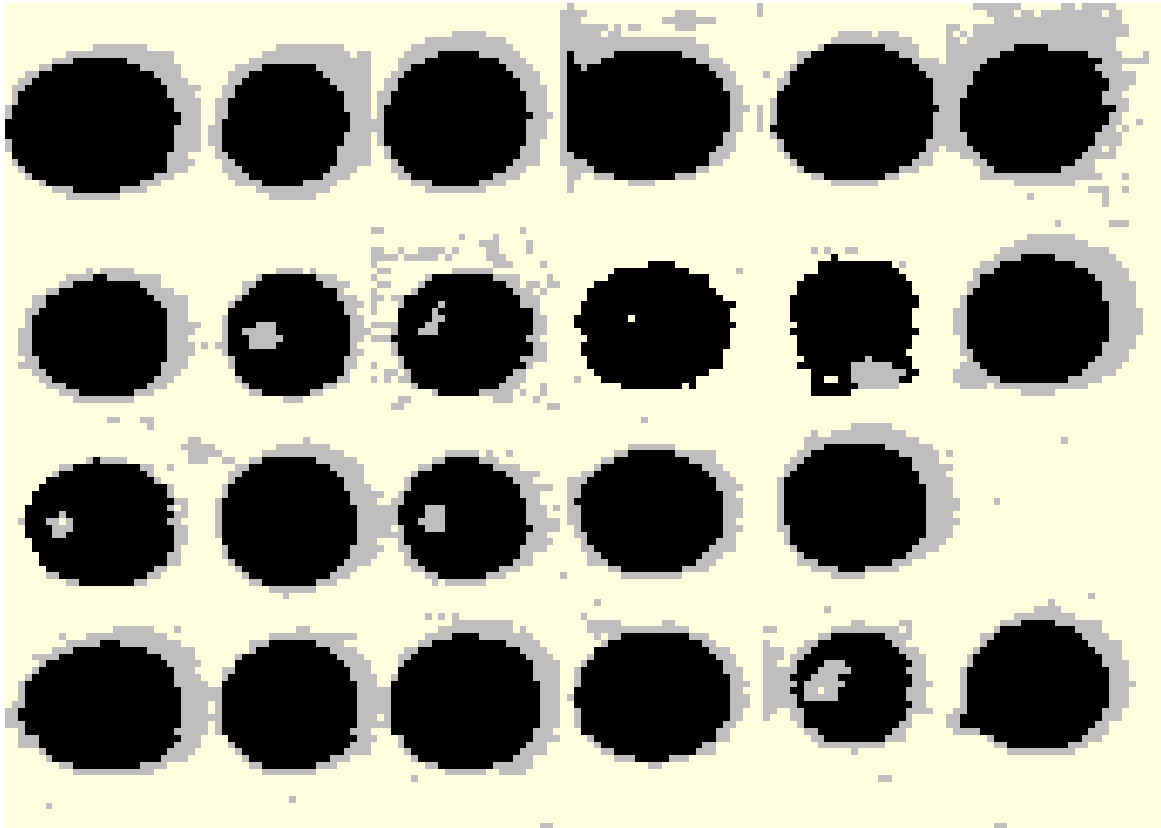


Figure 4: Segmented spots produced by `spotseg` for the HIV data subset, using the grid produced by `spotgrid`. The color scheme is as follows: *black* denotes the spots, *yellow* denotes background, *gray* denotes pixels of uncertain classification.

It is possible to process a subset of the regions in the grid using the arguments `R` for grid (as opposed to pixel) row location of the spot and `C` for grid column location. The `show` option in `spotseg` can be used to display details for each spot as it is classified. When more than one spot is processed, the graphics command `par(ask = TRUE)` should be set so that the displays can be stepped through. The following is an example of the segmenting and display of an individual spot.

```
hivSeg14 <- spotseg( chan1, chan2, hivGrid$rowcut, hivGrid$colcut,  
                    R = 1, C = 4, show = TRUE)
```

The resulting display is shown in Figure 5

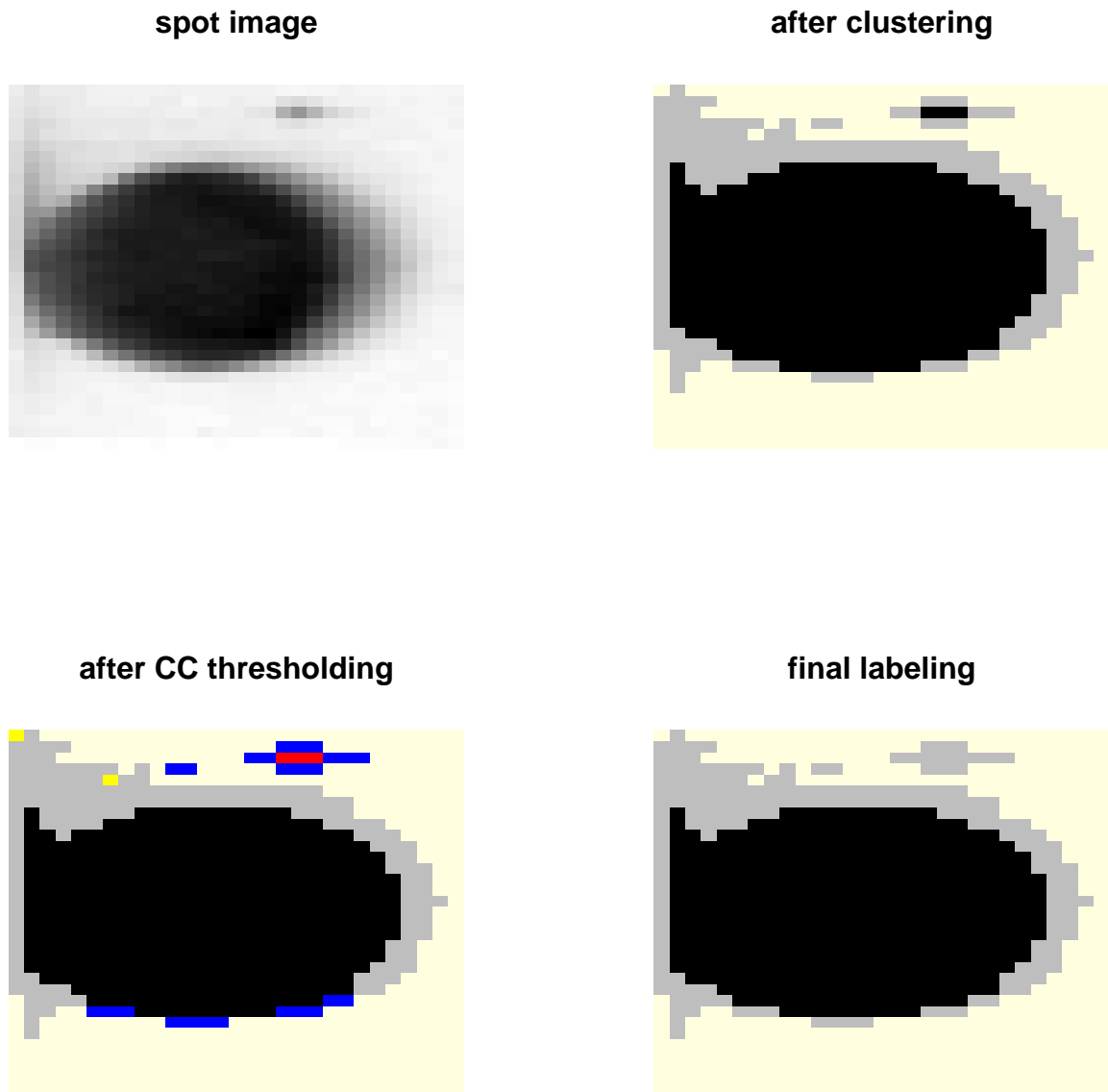


Figure 5: **spotseg** processing of the 1,4 section of the gridded HIV image subset. Clockwise from top left: gray-scale image, labeled image after model-based clustering (light yellow: lowest intensity; black: highest intensity), clustered image with connected components less than threshold in size labeled (bright yellow, blue, red denote components below threshold in size for the light yellow, gray, and black groups, respectively), final labeling.

Mean and median pixel intensities for the foreground and background for each channel and each spot can be recovered through the `summary` function applied to the output of `spotseg`. For example, the following extracts the summary intensities for the spot shown in Figure 5

```
> spotSumry <- summary(Seg)

> spotSumry$channel1$foreground$mean[1,4]
[1] 3693.215
> spotSumry$channel1$foreground$median[1,4]
[1] 4124.107
> spotSumry$channel2$foreground$mean[1,4]
[1] 7711.921
> spotSumry$channel2$foreground$median[1,4]
[1] 9479.41

> spotSumry$channel1$background$mean[1,4]
[1] 884.8382
> spotSumry$channel1$background$median[1,4]
[1] 877.9754
> spotSumry$channel2$background$mean[1,4]
[1] 273.5498
> spotSumry$channel2$background$median[1,4]
[1] 260.6310
```

References

- [1] C. Fraley and A. E. Raftery. MCLUST: Software for model-based cluster analysis. *Journal of Classification*, 16:297–306, 1999.
- [2] C. Fraley and A. E. Raftery. Enhanced software for model-based clustering, density estimation, and discriminant analysis: MCLUST. *Journal of Classification*, 20:263–286, 2003.
- [3] C. Fraley and A. E. Raftery. MCLUST version 3 for R: Normal mixture modeling and model-based clustering. Technical Report 504, University of Washington, Department of Statistics, September 2006.
- [4] Q. Li, C. Fraley, R. E. Bumgarner, K. Y. Yeung, and A. E. Raftery. Donuts, scratches, and blanks: Robust model-based segmentation of microarray images. *Bioinformatics*, 21:2875–2882, 2005.
- [5] A. B. van’t Wout, G. K. Lehrman, S. A. Mikeeva, G. C. O’Keefe, M. G. Katze, R. E. Bumgarner, G. K. Geiss, and J. I. Mullins. Cellular gene expression upon human immunodeficiency type 1 infection of CD4(+)-T-cell lines. *Journal of Virology*, 77(2):1392–1402, January 2003.