

A Graphical Tool for the Detection of Modes in Continuous Data

Thomas Burger^{1*}, Thierry Dhorne¹

1. Université Européenne de Bretagne, Université de Bretagne-Sud, CNRS, Lab-STICC, Centre de Recherche Yves Coppens BP 573, F-56017 Vannes cedex, France

* Contact author: thomas.burger@univ-ubs.fr

Keywords: mode estimation, kernel transform, multi-scale mean shift, dendrogram

In (Bickel, 2003) is presented a robust parametric estimator for the mode of a monomodal continuous distribution. Therefore, it is necessary that the distribution is monomodal. On the other hand, there have been some non-parametric methods for the estimation of the local modes of multimodal distributions. Here, we present a graphical tool that conveniently helps deciding on visual bases, the number of modes of a distribution.

To do so, the distribution is convoluted by a kernel of various scales to let local maxima of the density appear. Conceptually, the approach is similar to time-frequency analysis or wavelet analysis, but in order to best describe the shape of the distribution, Gaussian kernels are used. They are known to be more efficient in computer vision and pattern classification, and the corresponding representation fits the theoretical expectations (Mokhtarian, 1992).

Some other works have explored this connection between pattern classification and descriptive statistics. Hence, a work with ideas similar to ours has already been proposed to publication (Griffin, unpublished), but to our knowledge, in spite of its quality, it remains unpublished. It is based on a multi-scale mean shift algorithm, and the approach is once again rather formal: the point is more to find the various modes, than to provide a convenient way to represent them. Hence, in spite of a common theoretical framework (the similarity with time-frequency analysis in computer vision), the objective is somewhat different.

In addition to this work, we propose a dendrogram-like representation that helps the expert to describe the datasets and/or to propose an adapted mixture model. From an experimental point of view, the method is validated on real and simulated datasets. Finally, an efficient implementation is given.

References

- BICKEL, D. (2003). Robust and efficient estimation of the mode of continuous data: The mode as a viable measure of central tendency, *Journal of statistical computation and simulation*, vol. 73, Issue 12, pp. 899-912.
- GRIFFIN, L. D., LILHOLM, M. (unpublished). A Multiscale Mean Shift Algorithm for Mode Estimation. Submitted in 2005 to *IEEE Transacton on Pattern Analysis Machine Intelligence*.
- MOKHTARIAN, F. and MACKWORTH, A. K.(1992). A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, Issue 8, pp. 789-805.