Reviewer's report

Title: Application of Mean-Shift Clustering to Blood Oxygen Level Dependent Functional MRI Activation Detection

Version: 2  Date: 5 November 2013

Reviewer: Sebastian Domsch

Reviewer's report:

The goal of the present study was to improve the detectability of blood oxygen level dependent (BOLD) activation in function Magnetic Resonance Imaging (fMRI) by taking advantages of relationships of neighbouring voxels which is referred to as the mean shift clustering (MSC) technique. The MSC technique, originally used for pattern recognition, image processing, and image segmentation, was applied to simulated and real fMRI datasets and compared to existing methods used for the detection of BOLD activation such as a cross correlation analysis (CCA) method and a cluster analysis (CA) method. Unfortunately, the results of this comparison are presented very poorly. First of all, the simulation part brings up many unresolved questions. But the major concern is that the results of the real fMRI datasets are only qualitatively shown for a single slice in a single dataset. Apparently ten subjects were measure but for whatever reasons the single subject data were averaged prior to statistical analysis. Moreover, neither the number of activated voxel nor maximal t-values or any other quantitative measure was shown. Last but not least, the three different methods were compared at individual statistical thresholds making a fair comparison impossible.

Major Issues

1. The simulation part makes seems appropriate but the description is incomplete. It is not clear, why the false positive rate actually decreases when the statistical parameter image (SPI) from the CCA is used as input image for the MSC algorithm. How the MSC method generally works is well described (except for the mathematical derivation) but not how falsely activated voxels in the SPI are recognized as such. One can only speculate that isolated voxel, for which no cluster can be identified, are considered as false positives. Please derive Eq. 3 in more mathematical detail and explain how false-positive in the SPI are recognized.

2. Real fMRI datasets were acquired from ten healthy subjects and averaged prior to statistical analysis. By doing so, much of the potential information about the strength and weaknesses of the MSC method are lost. The authors should present a table with individual fMRI results showing at least t-values and number of activated voxels for brain areas with task related BOLD activity.

3. The authors compare a standard cross correlation analysis (CCA) to a CCA+Clusteranalysis, and to a CCA+MSC. However, the results are compared
at different statistical thresholds which make a fair comparison impossible. The authors claim that the methods can’t be compared at a fixed threshold. If this is true, the authors should explain explicitly why this is the case. Further, they should make a workaround e.g. show maximal t-values or t-value distributions not depending on an arbitrary threshold.

4. The statistical analysis was performed for spatially filtered and unfiltered datasets. To show the unfiltered dataset is not of major importance since spatial smoothing is a standard part of post-processing in fMRI as recognized by the authors. The authors should skip the SNR comparison or explain in more detail why this is important to show.

5. For the CCA results, the statistical threshold (Z=4.8) was obviously chosen too high since no activation was detected in the M1/S1 brain region at that threshold. Those results were compared to the results using the standard cluster analysis (CA) at a threshold of Z=2.6 and to those using the MSC method at a threshold of Z=2. I claim that using the same threshold of e.g. Z=2.6 for the CCA datasets would yield approx. the same activation as detected in the CCA+CA and in the CCA+MSC datasets. The authors should choose a threshold for which BOLD activation can also be detected in the standard CCA datasets. By doing this, one could judge, if the detection of fine-scale activation pattern are actually improved using the proposed MSC method as claimed by the authors. For this purpose, images with nice activation pattern could be convincing even though there is no real measure since there is no ground truth. This should be discussed.

6. Last but not least, a third dataset is presented which was acquired during an fMRI resting state experiment, which was introduced completely unmotivated. The authors then inserted simulated activation and called it the “half simulated resting state dataset”. To me, this dataset is completely needless to show. In case it is not, the authors should provide a convincing argument in the Material section explaining why this dataset is important to be acquired i.e. what can be investigated using this half-simulated dataset what cannot be investigated using the simulated or the real fMRI datasets.

7. The authors stumble across two interesting points namely the dependence of the activation on the kernel size of the MSC algorithm and the noise characteristics of fMRI signals. It shows that depending on the kernel size used, the BOLD activation is either quite robust or it disappears completely. I would suggest to investigate this issue in more detail e.g. show single subject results (tmax, # activated voxels) using different kernel sizes. Moreover, the authors asse that the noise in real fMRI datasets is different from the Gaussian white noise assumed in the simulation. One might use different noise types in the simulation part and analyse its effects on true and false positive rates.

8. The authors also claim that the proposed method improves the detection of small and highly focused activation since their method does not require a cluster size threshold as other standard cluster analysis methods. To me this is not intuitive. The authors should either provide evidence to support this thesis or not mention it.

9. As correctly stated by the authors, it is important to find an appropriate kernel
size for the MSC algorithm. However, this question is not solved and the reader can only guess which kernel size is useful. The kernel size was varied for the simulated datasets but not for the real and therefore most important datasets. Since the simulation does not represent reality, as correctly recognised by the authors, the kernel size should also be varied for the real fMRI datasets which would be highly interesting.

10. Except for the well-written abstract and introduction, the text contains a lot of uncommon writing and redundancies. E.g. “It has one hundred images…” or “BOLD fMRI data was recorded. The BOLD data was acquired…” or “While the calculations performed in the MSC technique itself does not need to be modified for it to be adopted for use in fMRI, …” or “dramatic improvement” and many more. Please correct all instances.

11. The SNR of the simulated dataset was 0.3 and 0.5? So it contained more noise than signal? Please clarify.

Minor Issues

1. Please insert page and line number for easier review purposes.
2. Not figure xx but Figure xx. Please correct all instances.
3. Extremely bad writing in the results section “ROC Comparison of different kernel sizes”. Please correct.
4. In the figure legend is not supposed to explain results. Please correct all instances.
5. In Fig. 3 the authors speak of a dramatic (probably mean significant) improvement in the 20x20 activation pattern. Please show.
6. In Fig. 1, please also show the true positive rate for the standard CCA method so the reader can notice whether there is an improvement or not.
7. Show some SPI maps for the simulated datasets before and after applying the MSC method. This would add credibility to the simulation results.
8. Generally, the results should be explained in a lot more detail. What causes the peaks in Fig.1 etc. Why seem the true positive rates independent from the cluster sizes? These questions should at least be discussed.
9. Figure 2: How is it that the false positive rates are the same for the high and the low SNR dataset? Please explain.
10. Figure 2: Why does the false positive rate in the MSC datasets not increase to 100% when the z-score is decreasing? For a z-score of zero, all voxels should be activated. Hence, the number of false positives naturally increases. Please clarify!
11. For very small z-scores in the SPIs, the false positive rates significantly increase (see Fig.2). If I am correct, those SPI are further processed using the proposed MSC algorithm. To me it remains unclear how the MSC algorithm is able to differentiate clusters if all voxels in the SPI are activated. Please explain.
12. Legend of Fig.2: CCA for which SNR? Please give information in the legend.
13. Generally, all figure legend should be more informative. Please correct.

**Level of interest:** An article of importance in its field

**Quality of written English:** Not suitable for publication unless extensively edited

**Statistical review:** No, the manuscript does not need to be seen by a statistician.