Conclusion

This thesis deals with a new and rather general multilevel approach for cluster analysis in high-dimensional data. In contrast to known cluster methods it applies not only for geometric, but also for dynamic cluster problems.

To guarantee the applicability to large cluster problems, the cluster identification is done via a decomposition based representative clustering method. If the underlying decomposition is fine enough, this method allows a problem reduction without destroying the original cluster structure. Furthermore, an efficient cluster description becomes possible if one uses a special decomposition variant, called approximate box decomposition. The computation of a suitably fine decomposition is done by a self-organized neural network.

Upon using the theory of Perron Cluster analysis, the general multilevel cluster approach can been extended: For cluster problems with a stochastic homogeneity function it allows to compute a correct set of clusters, even if their number is unknown a priori. Since traditional cluster methods need the number of clusters as an input, this is a significant improvement. Furthermore, the extended approach allows for the first time a conformational analysis of large biomolecules in combination with hierarchical temperature embedding.

On the computational complexity side, the computation of a suitably fine decomposition is still the *bottleneck*. Especially for an application within commercial virtual screening projects, a speed-up will be necessary. In this respect, parallelization and an improved convergence of the SOM/SOBM algorithm seem to be promising.