Chapter 3

Application to a virtual set of structures

The validity and applicability of the last two steps (characterization and classification) of the proposed methodology were tested on a virtual set of nearly 200 surfaces created using computer three-dimensional modeling tools. The extraction step was not included in this test, since the starting point is the set of surfaces itself. Nevertheless, modeled structures of very different sizes were included, to emulate the multiple scales that would result from the extraction step, had it been included. Also, the shapes of the modeled structures are all different. They could be visually classified into three main groups with a common geometry: blobs, tubes, and sheets. The target of this test was to educe those three main groups automatically and without any a priori knowledge of the possible geometries of the structures present in the dataset or of the number of groups among them, that is, simply based on the characterization and classification steps of the methodology previously explained. Among the modeled sheet-like structures, approximately one third were given a certain rolling geometry (spiral-like sheets).

Figure 3.1 shows the visualization space with the results of the test. Each sphere in that space represents a structure of the virtual set (some examples are projected onto the planar sides). The color of each sphere in the visualization space indicates the cluster to which its corresponding structure has been automatically assigned by the clustering algorithm during the classification step of the methodology, and its diameter is scaled using its associated silhouette coefficient, which represents for each structure the degree of membership to the cluster to which it was assigned (refer

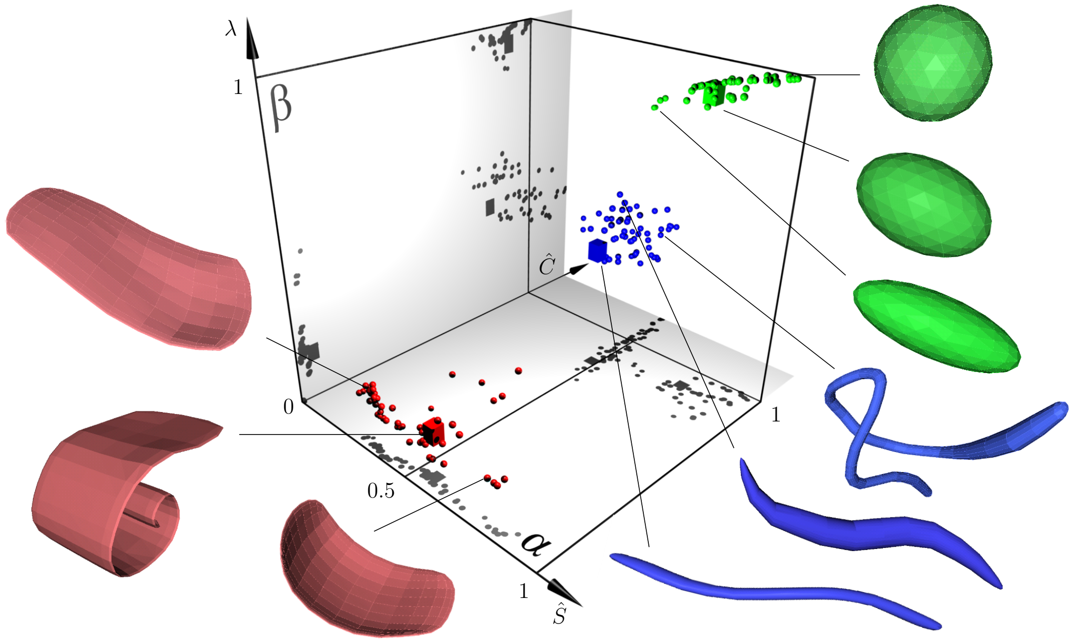

Figure 3.1: Visualization space with clustering results for the virtual set of modeled structures, with representative examples shown at the sides

to §2.3.3), renormalized to have positive values that allow a comparison among structures. For reference, the closest elements to the cluster centers have been highlighted using cubes of slightly bigger size.

Three clusters were automatically educed in the classification step and each structure was 'correctly' assigned by the algorithm to the appropriate group corresponding to the previously constructed geometry. This can be seen in Figure 3.1 from the relative locations of the centers of the glyphs defined by $\hat{S}$, $\hat{C}$, and λ for each structure. We emphasize that neither information on the previously constructed shapes nor the number of groups to be educed formed any part of the clustering algorithm (for example as pre-conditioning). The results of Figure 3.1 are a consequence of the geometric characterization and automatic classification in the feature space of parameters.

We note also that in Figure 3.1 (as was sketched in Figure 2.8), the sheet-like structures can spread over a large region near the plane $\hat{C} = 0$ in the visualization space. This region could be narrowed by means of a transformation of the $(\hat{S}, \hat{C})$-plane to Cartesian coordinates ($\hat{X} = \hat{C}\cos[\pi(\hat{S} - 1/2)], \hat{Y} = \hat{C}\sin[\pi(\hat{S} - 1/2)]$). This would bring sheet-like structures to the axis $(0, 0, \lambda)$ in the new visualization space. Nevertheless, it is helpful to keep the original visualization and feature spaces, since that allows a possible distinction of the different shapes of the structures that fall into the broadly defined sheet-like geometry. For example, in the test case of the virtual set of modeled structures presented here, the second optimum automatic clustering result was such that four groups were educed: the blob-like and tube-like clusters remained the same as in the optimum case of three clusters described above, but the sheet-like cluster was split into two, with one of these clusters containing a large proportion of structures with a rolling geometry (spiral-like sheets). This also suggests that further post-processing (ideally also automatic) of the educed clusters can be helpful in refining the results.