Hashing for similarity search is one of the most widely used methods to solve the approximate nearest neighbor search problem. In this method, one first maps data items from a real valued high-dimensional space to a suitable low dimensional binary code space and then performs the approximate nearest neighbor search in this code space instead. This is beneficial because the search in the code space can be solved more efficiently in terms of runtime complexity and storage consumption. Obviously, for this method to succeed, it is necessary that similar data items be mapped to binary code words that have small Hamming distance.

For real-world data such as images, one usually proceeds as follows. For each data item, a pre-processing algorithm removes noise and insignificant information and extracts important discriminating information to generate a feature vector that captures the important semantic content. Next, a vector hash function maps this real valued feature vector to a binary code word. It is also possible to use the raw feature vectors afterwards to further process the search result candidates produced by binary hash codes.

In this dissertation we focus on the following. First, developing a learning based counterpart for the MinHash hashing algorithm.

Second, presenting a new unsupervised hashing method UmapHash to map the neighborhood relations of data items from the feature vector space to the binary hash code space. Finally, an application of the aforementioned hashing methods for rapid face image recognition.