Vision, Modeling, and Visualization (2014)
J. Bender, A. Kuijper, T. von Landesberger, H. Theisel and P. Urban (Eds.)

Hierarchical Hashing for Pattern Search in 3D Vector Fields
Additional Material

Zhongjie Wang[1], Hans Peter Seidel[1], Tino Weinkauf[2]

[1]MPI Informatics, Fraunhofer IGD, Saarbrücken, Germany
[2]KTH Royal Institute of Technology, Stockholm, Sweden

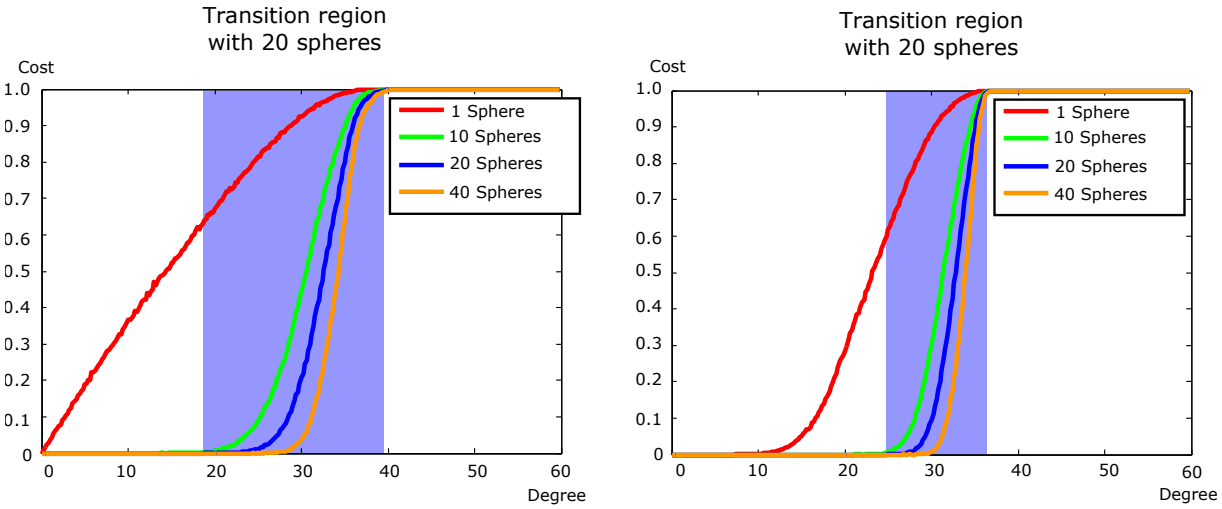

Figure 1: Accuracy curves of different LSH strategies and different number of hashing functions (Spheres). Left: classic LSH algorithm. Right: Multi-probe LSH algorithm.

Abstract

*In this material, we present an analysis of different LSH strategies and parameters. Firstly, We evaluate the accuracy of hashing retrieval under different number of hashing functions. Secondly, we compare the classic LSH algorithm [IM98] and multi-probe LSH algorithm [LJW*07] based on their accuracy of retrieval. Our selections of strategy and parameters are all based on these evaluations.*

1. Evaluation of Retrieval Accuracy

We conduct an experiment to perform this evaluation. Given two unit vectors **a** and **b** with angle α, we generate 20000 random vector pairs by transforming **a** and **b** in combination with random rotation matrices. In any LSH algorithm, If **a** and **b** are considered similar, idealy all the transformed vector pairs with angle α should be considered similar. In this experiment, we compute the dissimilarity (cost) of two vectors with angle α as the percentage of all 20000 pairs which are considered as different. We repeat this computation for the angles from 0.0 to 60.0. The resulting curve describes the accuracy of the retrieval algorithm. In perfection, the accuracy curve should be a step function, i.e., starting from 0, jumping to 1 at middle and keeping the value to the end, while in practice, a high accuracy means the transition region where the curve jumps from 0 to 1 should be narrow. In Figure 1, we plot the curves under different number of hashing functions (Spheres). In both algorithms, we observe that the more hashing functions used, the better accuracy obtained. To ballance

the accuracy and efficiency, we choose 20 hashing functions in our algorithm.

2. Comparison of LSH strategies

We compare the classic LSH algorithm and multi-probe LSH algorithm also using the accuracy curves as mentioned above. In Figure 1, we observe that when both algorithms use 20 hashing functions, the transition region of multi-probe LSH is much smaller than that of classic LSH algorithm. Based on this observation, we use multi-probe LSH algorithm in this paper.

References

[IM98] INDYK P., MOTWANI R.: Approximate nearest neighbors: towards removing the curse of dimensionality. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing* (1998), ACM, pp. 604–613. 1

[LJW*07] LV Q., JOSEPHSON W., WANG Z., CHARIKAR M., LI K.: Multi-probe lsh: efficient indexing for high-dimensional similarity search. In *Proceedings of the 33rd international conference on Very large data bases* (2007), VLDB Endowment, pp. 950–961. 1