Fast Image Search for Learned Metrics

Prateek Jain, Brian Kulis, and Kristen Grauman

University of Texas at Austin
1 University Station C0500
Taylor Hall 2.124
University of Texas at Austin
Austin, TX 78712-1188

{pjain, kulis, grauman}@cs.utexas.edu
http://www.cs.utexas.edu/~grauman/

As the world's store of digital images continues to grow exponentially, and as novel data-rich approaches to computer vision begin to emerge, many interesting problems demand fast techniques capable of accurately searching very large databases of images or image features. For most such tasks, the quality of the results relies heavily on the chosen image representation and the distance metric used to compare examples. Unfortunately, preferred representations tend to be high-dimensional [4,5], and often the best distance metric is one specialized (or learned) for the task at hand [1,2,3,7], rather than, say, a generic Euclidean norm or Gaussian kernel. Neither factor bodes well for large-scale image search: known data structures for efficient exact search are ineffective for high-dimensional spaces, while existing methods for approximate sub-linear time search are defined only for certain standard metrics.

In this work we develop a general algorithm that enables fast approximate similarity search for a family of *learned* metrics and kernel functions. Given pairwise similarity and dissimilarity constraints between some images, we learn a Mahalanobis distance function that captures the images' underlying relationships well. To allow sub-linear time similarity search under the learned metric, we show how to encode the learned metric parameterization into randomized locality-sensitive hash functions. We further formulate an indirect solution that enables metric learning and hashing for vector spaces whose high dimensionality makes it infeasible to learn an explicit weighting over the feature dimensions.

While randomized algorithms such as locality-sensitive hashing (LSH) have frequently been employed to mitigate the time complexity of identifying similar examples—particularly in vision [6]—their use has been restricted to generic measures for which the appropriate hash functions are already defined; that is, direct application to learned metrics was not possible. We instead devise a method that allows knowledge attained from partially labeled data or paired constraints to be incorporated into the hash functions. Our algorithm is theoretically sound: there is provably no additional loss in accuracy relative to the learned metric beyond the quantifiable loss induced by the approximate search technique.

We demonstrate the generality of our approach by applying it to three distinct large-scale image search problems: exemplar-based recognition, pose estimation, and feature indexing. Our method allows rapid and accurate retrieval, and gains over relevant state-of-the-art techniques.

References

[1] A. Frome, Y. Singer, F. Sha, and J. Malik. Learning Globally-Consistent Local Distance Functions for Shape-Based Image Retrieval and Classification. In *Proceedings of the IEEE International Conference on Computer Vision* (ICCV), Rio de Janeiro, Brazil, October 2007.

[2] K. Grauman and T. Darrell. The Pyramid Match Kernel: Discriminative Classification with Sets of Image Features. In *Proceedings of the IEEE International Conference on Computer Vision* (ICCV), Beijing, China, October 2005.

[3] T. Hertz, A. Bar-Hillel, and D. Weinshall. Learning Distance Functions for Image Retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Washington, DC, June 2004.

[4] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 2004.

[5] J. Sivic and A. Zisserman. Video Data Mining Using Configurations of Viewpoint Invariant Regions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Washington, D.C., June 2004.

[6] G. Shakhnarovich, T. Darrell, and P. Indyk, editors. *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice*. MIT Press, March 2006.

[7] K. Weinberger, J. Blitzer, and L. Saul. Distance Metric Learning for Large Margin Nearest Neighbor Classification. In *Advances in Neural Information Processing Systems* (NIPS), 2006.

Topic: learning algorithms, visual processing and pattern recognition
Preference: oral