

VISUAL RECOGNITION IN LARGE COLLECTIONS
REFERENCES OF THE TUTORIAL

Hervé Jégou

April 1st 2014**Part I**

- [Chandrasekhar et al., 2011] Chandrasekhar, V., Chen, D., Tsai, S., Cheung, N., Chen, H., Takacs, G., Reznik, Y., Vedantham, R., Grzeszczuk, R., Bach, J., and Girod, B. (2011). The stanford mobile visual search dataset. In *MMSys*.
- [Douze et al., 2009] Douze, M., Jégou, H., Singh, H., Amsaleg, L., and Schmid, C. (2009). Evaluation of GIST descriptors for web-scale image search. In *CIVR*.
- [Flickner et al., 1995] Flickner, M., Harpreet, S., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., and Yanker, P. (1995). Query by image and video content: the qbic system. *Computer*, 28:23–21.
- [Jégou et al., 2008] Jégou, H., Douze, M., and Schmid, C. (2008). Hamming embedding and weak geometric consistency for large scale image search. In *ECCV*.
- [Jégou et al., 2010] Jégou, H., Douze, M., Schmid, C., and Pérez, P. (2010). Aggregating local descriptors into a compact image representation. In *CVPR*.
- [Jégou et al., 2012] Jégou, H., Perronnin, F., Douze, M., Sánchez, J., Pérez, P., and Schmid, C. (2012). Aggregating local descriptors into compact codes. In *IEEE Trans. PAMI*.
- [Nistér and Stewénius, 2006] Nistér, D. and Stewénius, H. (2006). Scalable recognition with a vocabulary tree. In *CVPR*, pages 2161–2168.
- [Philbin et al., 2007] Philbin, J., Chum, O., Isard, M., Sivic, J., and Zisserman, A. (2007). Object retrieval with large vocabularies and fast spatial matching. In *CVPR*.
- [Quack et al., 2004] Quack, T., Mönich, U., Thiele, L., and Manjunath, B. (2004). Cortina: A system for large-scale, content-based web image retrieval. In *ACM Multimedia*.
- [Sivic and Zisserman, 2003] Sivic, J. and Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. In *ICCV*, pages 1470–1477.
- [Torralba et al., 2008] Torralba, A., Fergus, R., and Weiss, Y. (2008). Small codes and large databases for recognition. In *CVPR*.

Part II

- [Arandjelovic and Zisserman, 2012] Arandjelovic, R. and Zisserman, A. (2012). Three things everyone should know to improve object retrieval. In *CVPR*.
- [Bay et al., 2008] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (SURF). *CVIU*, 110(3):346–359.
- [Delhumeau et al., 2013] Delhumeau, J., Gosselin, P.-H., Jégou, H., and Pérez, P. (2013). Re-visiting the vlad image representation. In *ACM Multimedia*.
- [Douze et al., 2009] Douze, M., Jégou, H., Singh, H., Amsaleg, L., and Schmid, C. (2009). Evaluation of GIST descriptors for web-scale image search. In *CIVR*.
- [Fergus et al., 2005] Fergus, R., Fei-Fei, L., Perona, P., and Zisserman, A. (2005). Learning object categories from google’s image search. In *ICCV*.
- [Flickner et al., 1995] Flickner, M., Harpreet, S., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., and Yanker, P. (1995). Query by image and video content: the qbic system. *Computer*, 28:23–21.
- [Gordo et al., 2012] Gordo, A., Rodriguez-Serrano, J. A., Perronnin, F., and Valveny, E. (2012). Leveraging category-level labels for instance-level image retrieval. In *CVPR*.
- [Jain et al., 2012] Jain, M., Benmokhtar, R., Gros, P., and Jégou, H. (2012). Hamming embedding similarity-based image classification. In *ICMR*.
- [Lazebnik et al., 2006] Lazebnik, S., Schmid, C., and Ponce, J. (2006). Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In *CVPR*.
- [Leung and Malik, 1999] Leung, T. and Malik, J. (1999). Recognizing surfaces using three-dimensional textons. In *CVPR*.
- [Leung and Malik, 2001] Leung, T. and Malik, J. (2001). Representing and recognizing the visual appearance of materials using three-dimensional textons. *IJCV*, 43(1):29–44.
- [Lowe, 2004] Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *IJCV*.
- [Lowe, 1999] Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *ICCV*, pages 1150–1157.
- [Matas et al., 2002] Matas, J., Chum, O., Martin, U., and Pajdla, T. (2002). Robust wide baseline stereo from maximally stable extremal regions. In *BMVC*, pages 384–393.
- [Mikolajczyk and Schmid, 2004] Mikolajczyk, K. and Schmid, C. (2004). scale and affine invariant interest point detectors. *IJCV*, 60(1):63–86.
- [Oliva and Torralba, 2001] Oliva, A. and Torralba, A. (2001). Modeling the shape of the scene: a holistic representation of the spatial envelope. *IJCV*, 42(3):145–175.
- [Schmid and Mohr, 1997] Schmid, C. and Mohr, R. (1997). Local grayvalue invariants for image retrieval. *IEEE Trans. PAMI*, 19(5):530–535.
- [Swain and Ballard, 1991] Swain, M. J. and Ballard, D. H. (1991). Color indexing. *IJCV*, 7:11–32.
- [Winder and Brown, 2007] Winder, S. and Brown, M. (2007). Learning local image descriptors. In *CVPR*.

Part III

- [Arandjelovic and Zisserman, 2012] Arandjelovic, R. and Zisserman, A. (2012). Three things everyone should know to improve object retrieval. In *CVPR*.
- [Bo and Sminchisescu, 2009] Bo, L. and Sminchisescu, C. (2009). Efficient match kernels between sets of features for visual recognition. In *NIPS*.
- [Broder, 1997] Broder, A. Z. (1997). On the resemblance and containment of documents. In *Proc. Compression and Complexity of Sequences*.
- [Chum and Matas, 2010] Chum, O. and Matas, J. (2010). Large-scale discovery of spatially related images. *IEEE Trans. PAMI*, 32(2):371–377.
- [Chum et al., 2011] Chum, O., Mikulik, A., Perdoch, M., and Matas, J. (2011). Total recall II: Query expansion revisited. In *CVPR*.
- [Chum et al., 2009] Chum, O., Perdoch, M., and Matas, J. (2009). Geometric min-hashing: Finding a (thick) needle in a haystack. In *CVPR*.
- [Chum et al., 2007] Chum, O., Philbin, J., Sivic, J., Isard, M., and Zisserman, A. (2007). Total recall: Automatic query expansion with a generative feature model for object retrieval. In *ICCV*.
- [Csurka et al., 2004] Csurka, G., Dance, C., Fan, L., Willamowski, J., and Bray, C. (2004). Visual categorization with bags of keypoints. In *ECCV Workshop Statistical Learning in Computer Vision*.
- [Duda and Hart, 1972] Duda, R. R. and Hart, P. E. (1972). Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15:11–15.
- [Hough, 1962] Hough, P. V. C. (1962). Method and means for recognizing complex patterns. US patent 3069654.
- [Jégou et al., 2008] Jégou, H., Douze, M., and Schmid, C. (2008). Hamming embedding and weak geometric consistency for large scale image search. In *ECCV*.
- [Jégou et al., 2009a] Jégou, H., Douze, M., and Schmid, C. (2009a). On the burstiness of visual elements. In *CVPR*.
- [Jégou et al., 2009b] Jégou, H., Douze, M., and Schmid, C. (2009b). Packing bag-of-features. In *ICCV*.
- [Jégou et al., 2007] Jégou, H., Harzallah, H., and Schmid, C. (2007). A contextual dissimilarity measure for accurate and efficient image search. In *CVPR*.
- [Lazebnik et al., 2006] Lazebnik, S., Schmid, C., and Ponce, J. (2006). Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In *CVPR*.
- [Leung and Malik, 1999] Leung, T. and Malik, J. (1999). Recognizing surfaces using three-dimensional textons. In *CVPR*.
- [Lin and Brandt, 2010] Lin, Z. and Brandt, J. (2010). A local bag-of-features model for large-scale object retrieval. *ECCV*.

- [Lowe, 2004] Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110.
- [Muja and Lowe, 2009] Muja, M. and Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. In *VISAPP*.
- [Nistér and Stewénus, 2006] Nistér, D. and Stewénus, H. (2006). Scalable recognition with a vocabulary tree. In *CVPR*, pages 2161–2168.
- [Paulevé et al., 2010] Paulevé, L., Jégou, H., and Amsaleg, L. (2010). Locality sensitive hashing: a comparison of hash function types and querying mechanisms. *Pattern Recognition Letters*, 31(11):1348–1358.
- [Perdoch et al., 2009] Perdoch, M., Chum, O., and Matas, J. (2009). Efficient representation of local geometry for large scale object retrieval. In *CVPR*.
- [Perronnin et al., 2010] Perronnin, F., J.Sánchez, and Mensink, T. (2010). Improving the fisher kernel for large-scale image classification. In *ECCV*.
- [Philbin et al., 2007] Philbin, J., Chum, O., Isard, M., Sivic, J., and Zisserman, A. (2007). Object retrieval with large vocabularies and fast spatial matching. In *CVPR*.
- [Philbin et al., 2008] Philbin, J., Chum, O., Isard, M., Sivic, J., and Zisserman, A. (2008). Lost in quantization: Improving particular object retrieval in large scale image databases. In *CVPR*.
- [Quack et al., 2007] Quack, T., Ferrari, V., Leibe, B., and Van Gool, L. (2007). Efficient mining of frequent and distinctive feature configurations. In *ICCV*, pages 1–8.
- [Robertson and arck Jones, 1976] Robertson, S. E. and arck Jones, K. S. (1976). Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27(3).
- [Romberg and Lienhart, 2013] Romberg, S. and Lienhart, R. (2013). Bundle min-hashing for logo recognition. In *ICMR*, pages 113–120.
- [Sivic and Zisserman, 2003] Sivic, J. and Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. In *ICCV*, pages 1470–1477.
- [Stewenius et al., 2012] Stewenius, H., Gunderson, S. H., and Pilet, J. (2012). Size matters: Exhaustive geometric verification for image retrieval. In *ECCV*.
- [Tuytelaars and Schmid, 2007] Tuytelaars, T. and Schmid, C. (2007). Vector quantizing feature space with a regular lattice. In *ICCV*.
- [Wu et al., 2009] Wu, Z., Ke, Q., Isard, M., and Sun, J. (2009). Bundling features for large scale partial-duplicate web image search. In *CVPR*, pages 25–32.
- [Zhang et al., 2008] Zhang, J., Long, X., and Suel, T. (2008). Performance of compressed inverted list caching in search engines. In *WWW*, pages 387–396.
- [Zhang et al., 2011] Zhang, Y., Jia, Z., and Chen, T. (2011). Image retrieval with geometry-preserving visual phrases. In *CVPR*, pages 809–816.
- [Zhao et al., 2013] Zhao, W.-L., Jégou, H., and Gravier, G. (2013). Sim-min-hash: An efficient matching technique for linking large image collections. In *ACM Multimedia*.

- [Zhao et al., 2010] Zhao, W.-L., Wu, X., and Ngo, C.-W. (2010). On the annotation of web videos by efficient near-duplicate search. *IEEE Trans. on Multimedia*, 12(5):448–461.
- [Zhu et al., 2013] Zhu, C.-Z., Jégou, H., and ichi Satoh, S. (2013). Query-adaptive asymmetrical dissimilarities for visual object retrieval. In *ICCV*.
- [Zobel and Moffat, 2006] Zobel, J. and Moffat, A. (2006). Inverted files for text search engines. *ACM Computing Surveys*, 38(2):6.

Part IV

- [Bo and Sminchisescu, 2009] Bo, L. and Sminchisescu, C. (2009). Efficient match kernels between sets of features for visual recognition. In *NIPS*.
- [Chatfield et al., 2011] Chatfield, K., Lempitsky, V., Vedaldi, A., and Zisserman, A. (2011). The devil is in the details: an evaluation of recent feature encoding methods. In *BMVC*.
- [Delhumeau et al., 2013] Delhumeau, J., Gosselin, P.-H., Jégou, H., and Pérez, P. (2013). Revisiting the vlad image representation. In *ACM Multimedia*.
- [Grauman and Leibe, 2011] Grauman, K. and Leibe, B. (2011). *Visual object recognition*. Number 11. Morgan & Claypool Publishers.
- [Jaakkola and Haussler, 1998] Jaakkola, T. and Haussler, D. (1998). Exploiting generative models in discriminative classifiers. In *NIPS*.
- [Jégou et al., 2010a] Jégou, H., Douze, M., and Schmid, C. (2010a). Improving bag-of-features for large scale image search. *IJCV*, 87(3):316–336.
- [Jégou et al., 2010b] Jégou, H., Douze, M., Schmid, C., and Pérez, P. (2010b). Aggregating local descriptors into a compact image representation. In *CVPR*.
- [Jégou et al., 2012] Jégou, H., Perronnin, F., Douze, M., Sánchez, J., Pérez, P., and Schmid, C. (2012). Aggregating local descriptors into compact codes. In *IEEE Trans. PAMI*.
- [Lyu, 2005] Lyu, S. (2005). Mercer kernels for object recognition with local features. In *CVPR*.
- [Perronnin and Dance, 2007] Perronnin, F. and Dance, C. R. (2007). Fisher kernels on visual vocabularies for image categorization. In *CVPR*.
- [Perronnin et al., 2010] Perronnin, F., Liu, Y., Sanchez, J., and Poirier, H. (2010). Large-scale image retrieval with compressed Fisher vectors. In *CVPR*.
- [Picard and Gosselin, 2011] Picard, D. and Gosselin, P.-H. (2011). Improving image similarity with vectors of locally aggregated tensors. In *ICIP*.
- [Tolias et al., 2013] Tolias, G., Avrithis, Y., and Jégou, H. (2013). To aggregate or not to aggregate: Selective match kernels for image search. In *ICCV*.
- [Torralba et al., 2008] Torralba, A., Fergus, R., and Weiss, Y. (2008). Small codes and large databases for recognition. In *CVPR*.
- [Zhou et al., 2010] Zhou, X., Yu, K., Zhang, T., and Huang, T. S. (2010). Image classification using super-vector coding of local image descriptors. In *Computer Vision–ECCV 2010*, pages 141–154.

Part V

- [Andoni and Indyk, 2006] Andoni, A. and Indyk, P. (2006). Near-optimal hashing algorithms for near neighbor problem in high dimensions. In *Proceedings of the Symposium on the Foundations of Computer Science*, pages 459–468.
- [Babenko and Lempitsky, 2012] Babenko, A. and Lempitsky, V. (2012). The inverted multi-index. In *CVPR*.
- [Belkin and Niyogi, 2003] Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396.
- [Bourrier et al., 2012] Bourrier, A., Perronnin, F., Gribonval, R., Pérez, P., and Jégou, H. (2012). Nearest neighbor search for arbitrary kernels with explicit embeddings. Research Report RR-8040, INRIA.
- [Charikar, 2002] Charikar, M. S. (2002). Similarity estimation techniques from rounding algorithms. In *STOC*, pages 380–388.
- [Datar et al., 2004] Datar, M., Immorlica, N., Indyk, P., and Mirrokni, V. (2004). Locality-sensitive hashing scheme based on p-stable distributions. In *Proceedings of the Symposium on Computational Geometry*, pages 253–262.
- [Dong et al., 2008a] Dong, W., Charikar, M., and Li, K. (2008a). Asymmetric distance estimation with sketches for similarity search in high-dimensional spaces. In *SIGIR*, pages 123–130.
- [Dong et al., 2008b] Dong, W., Wang, Z., Josephson, W., Charikar, M., and Li, K. (2008b). Modeling lsh for performance tuning. In *CIKM*.
- [Fuchs, 2011] Fuchs, J.-J. (2011). Spread representations. In *ASILOMAR*.
- [Gammeter et al., 2010] Gammeter, S., Gassmann, A., Bossard, L., Quack, T., and Van Gool, L. (2010). Server-side object recognition and client-side object tracking for mobile augmented reality. In *Computer Vision and Pattern Recognition Workshops*.
- [Ge et al., 2013] Ge, T., He, K., Ke, Q., and Sun, J. (2013). Optimized product quantization for approximate nearest neighbor search. In *CVPR*.
- [Gionis et al., 1999] Gionis, A., Indyk, P., and Motwani, R. (1999). Similarity search in high dimension via hashing. In *Proceedings of the International Conference on Very Large DataBases*, pages 518–529.
- [Gong and Lazebnik, 2011] Gong, Y. and Lazebnik, S. (2011). Iterative quantization: A procrustean approach to learning binary codes. In *CVPR*.
- [Gordo and Perronnin, 2011] Gordo, A. and Perronnin, F. (2011). Asymmetric distances for binary embeddings. In *CVPR*.
- [Gordo et al., 2012] Gordo, A., Rodriguez-Serrano, J. A., Perronnin, F., and Valveny, E. (2012). Leveraging category-level labels for instance-level image retrieval. In *CVPR*.
- [Harchaoui et al., 2012] Harchaoui, Z., Douze, M., Paulin, M., Dudik, M., and Malick, J. (2012). Large-scale image classification with trace-norm regularization. In *CVPR*.
- [Indyk and Motwani, 1998] Indyk, P. and Motwani, R. (1998). Approximate nearest neighbors: towards removing the curse of dimensionality. In *STOC*, pages 604–613.

- [Indyk and Naor, 2007] Indyk, P. and Naor, A. (2007). Nearest-neighbor-preserving embeddings. *ACM Transactions on Algorithms (TALG)*, 3.
- [Jégou et al., 2008] Jégou, H., Douze, M., and Schmid, C. (2008). Hamming embedding and weak geometric consistency for large scale image search. In *ECCV*.
- [Jégou et al., 2009] Jégou, H., Douze, M., and Schmid, C. (2009). Searching with quantization: approximate nearest neighbor search using short codes and distance estimators. Technical Report RR-7020, INRIA.
- [Jégou et al., 2010a] Jégou, H., Douze, M., and Schmid, C. (2010a). Improving bag-of-features for large scale image search. *IJCV*, 87(3):316–336.
- [Jégou et al., 2011] Jégou, H., Douze, M., and Schmid, C. (2011). Product quantization for nearest neighbor search. *IEEE Trans. PAMI*, 33(1):117–128.
- [Jégou et al., 2010b] Jégou, H., Douze, M., Schmid, C., and Pérez, P. (2010b). Aggregating local descriptors into a compact image representation. In *CVPR*.
- [Jégou et al., 2012] Jégou, H., Furon, T., and Fuchs, J.-J. (2012). Anti-sparse coding for approximate nearest neighbor search. In *ICASSP*.
- [Jégou et al., 2007] Jégou, H., Harzallah, H., and Schmid, C. (2007). A contextual dissimilarity measure for accurate and efficient image search. In *CVPR*.
- [Jégou et al., 2011] Jégou, H., Tavenard, R., Douze, M., and Amsaleg, L. (2011). Searching in one billion vectors: re-rank with source coding. In *ICASSP*.
- [Joly and Buisson, 2011] Joly, A. and Buisson, O. (2011). Random maximum margin hashing. In *CVPR*.
- [Kulis and Grauman, 2009] Kulis, B. and Grauman, K. (2009). Kernelized locality-sensitive hashing for scalable image search. In *ICCV*.
- [Lv et al., 2007] Lv, Q., Josephson, W., Wang, Z., Charikar, M., and Li, K. (2007). Multi-probe LSH: Efficient indexing for high-dimensional similarity search. In *Proceedings of the International Conference on Very Large DataBases*, pages 950–961.
- [Muja and Lowe, 2009] Muja, M. and Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. In *VISAPP*.
- [Nistér and Stewénius, 2006] Nistér, D. and Stewénius, H. (2006). Scalable recognition with a vocabulary tree. In *CVPR*, pages 2161–2168.
- [Norouzi and Fleet, 2013] Norouzi, M. and Fleet, D. (2013). Cartesian k-means. In *CVPR*.
- [Norouzi et al., 2012] Norouzi, M., Punjani, A., and Fleet, D. J. (2012). Fast search in hamming space with multi-index hashing. In *CVPR*.
- [Paulevé et al., 2010] Paulevé, L., Jégou, H., and Amsaleg, L. (2010). Locality sensitive hashing: a comparison of hash function types and querying mechanisms. *Pattern Recognition Letters*, 31(11):1348–1358.
- [Philbin et al., 2008] Philbin, J., Chum, O., Isard, M., Sivic, J., and Zisserman, A. (2008). Lost in quantization: Improving particular object retrieval in large scale image databases. In *CVPR*.

- [Raginsky and Lazebnik, 2010] Raginsky, M. and Lazebnik, S. (2010). Locality-sensitive binary codes from shift-invariant kernels. In *NIPS*.
- [Revaud et al., 2013] Revaud, J., Douze, M., Schmid, C., and Jégou, H. (2013). Event retrieval in large video collections with circulant temporal encoding. In *CVPR*.
- [Salakhutdinov and Hinton, 2007] Salakhutdinov, R. and Hinton, G. (2007). Semantic hashing. In *SIGIR workshop on Information Retrieval and applications of Graphical Models*.
- [Salakhutdinov and Hinton, 2009] Salakhutdinov, R. and Hinton, G. (2009). Semantic hashing. *International Journal of Approximate Reasoning*, 50(7):969–978.
- [Slaney et al., 2012] Slaney, M., Lifshits, Y., and He, J. (2012). Optimal parameters for locality-sensitive hashing. *Proceedings of the IEEE*, 100(9):2604–2623.
- [Uchida et al., 2011] Uchida, Y., Agrawal, M., and Sakazawa, S. (2011). Accurate content-based video copy detection with efficient feature indexing. In *ICMR*.
- [Vedaldi and Zisserman, 2012a] Vedaldi, A. and Zisserman, A. (2012a). Efficient additive kernels via explicit feature maps. *IEEE Trans. PAMI*, 34:480–492.
- [Vedaldi and Zisserman, 2012b] Vedaldi, A. and Zisserman, A. (2012b). Sparse kernel approximations for efficient classification and detection. In *CVPR*.
- [Wang et al., 2012] Wang, J., Kumar, S., and Chang, S.-F. (2012). Semi-supervised hashing for large scale search. *IEEE Trans. PAMI*, 6(12):1.
- [Weiss et al., 2009] Weiss, Y., Torralba, A., and Fergus, R. (2009). Spectral hashing. In *NIPS*.

Part VI

- [Bo et al., 2010] Bo, L., Ren, X., and Fox, D. (2010). Kernel descriptors for visual recognition. In *NIPS*.
- [Bo and Sminchisescu, 2009] Bo, L. and Sminchisescu, C. (2009). Efficient match kernels between sets of features for visual recognition. In *NIPS*.
- [Dong et al., 2011] Dong, W., Charikar, M., and Li, K. (2011). Efficient k-nearest neighbor graph construction for generic similarity measures. In *WWW*.
- [Jégou and Zisserman, 2014] Jégou, H. and Zisserman, A. (2014). Triangulation embedding and democratic kernels for image search. In *CVPR*.
- [Trevisiol et al., 2013] Trevisiol, M., Jégou, H., Delhumeau, J., and Gravier, G. (2013). Retrieving geo-location of videos with a divide & conquer hierarchical multimodal approach. In *ICMR*.
- [Vedaldi and Zisserman, 2012] Vedaldi, A. and Zisserman, A. (2012). Efficient additive kernels via explicit feature maps. *IEEE Trans. PAMI*, 34:480–492.