REFERENCES

[1]   J. Wagemans *et al.*, "A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization.," *Psychol. Bull.*, vol. 138, no. 6, p. 1172, 2012.

[2]   Z. Kourtzi, L. R. Betts, P. Sarkheil, and A. E. Welchman, "Distributed neural plasticity for shape learning in the human visual cortex," *PLoS Biol.*, vol. 3, no. 7, 2005.

[3]   P. J. Kellman and M. E. Arterberry, "Infant visual perception," *Handb. Child Psychol.*, vol. 2, 2007.

[4]   V. Ramachandran and S. Anstis, "Extrapolation of motion path in human visual perception.," *Vision Res.*, 1983.

[5]   D. Rose and R. Blake, "Motion perception: from phi to omega," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 353, no. 1371, pp. 967–980, 1998.

[6]   W. Hu, N. Xie, L. Li, X. Zeng, and S. Maybank, "A survey on visual content-based video indexing and retrieval," *Syst. Man Cybern. Part C Appl. Rev. IEEE Trans. On*, vol. 41, no. 6, pp. 797–819, 2011.

[7]   J. R. R. Uijlings, A. W. M. Smeulders, and R. J. H. Scha, "Real-Time Visual Concept Classification," *Multimed. IEEE Trans. On*, vol. 12, no. 7, pp. 665–681, Nov. 2010, doi: 10.1109/TMM.2010.2052027.

[8]   Yang Mingqiang, Kpalma K. Idiyo, and Ronsin Joseph, "A Survey of Shape Feature Extraction Techniques," *Pattern Recognit.*, pp. 43–90, Nov. 2008.

[9]   A. B. Benitez *et al.*, "Semantics of Multimedia in MPEG-7," in *Proceedings. International Conference on Image Processing*, 2002, vol. 1, p. 1.

[10]   L. Rossetto *et al.*, "IMOTION—a content-based video retrieval engine," in *International Conference on Multimedia Modeling*, 2015, pp. 255–260.

[11]   J. Lokoč, G. Kovalčík, T. Souček, J. Moravec, and P. Čech, "VIRET: A video retrieval tool for interactive known-item search," in *Proceedings of the 2019 on International Conference on Multimedia Retrieval*, 2019, pp. 177–181.

[12]   L. Cinque, G. Ciocca, S. Levialdi, A. Pellicano, and R. Schettini, "Color-based image retrieval using spatial-chromatic histograms," *Image Vis. Comput.*, vol. 19, no. 13, pp. 979–986, 2001.

[13]   R. Chakravarti and X. Meng, "A study of colour histogram based image retrieval," in *Information Technology: New Generations, 2009. ITNG'09. Sixth International Conference on*, 2009, pp. 1323–1328.

[14]   F. C. Heilbron, W. Barrios, V. Escorcia, and B. Ghanem, "Scc: Semantic context cascade for efficient action detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3175–3184.

[15]   Y. Nakajima and H. Saito, "Efficient object-oriented semantic mapping with object detector," *IEEE Access*, vol. 7, pp. 3206–3213, 2018.

[16]   Y. Nishizumi *et al.*, "FPGA implementation of object recognition processor for HDTV resolution video using sparse FIND feature," in *2017 IEEE International Workshop on Signal Processing Systems (SiPS)*, 2017, pp. 1–6.

[17]   A. Gorave, S. Misra, O. Padir, A. Patil, and K. Ladole, "Suspicious Activity Detection Using Live Video Analysis," in *Proceeding of International Conference on Computational Science and Applications*, 2020, pp. 203–214.

[18] H. Kiani Galoogahi, A. Fagg, C. Huang, D. Ramanan, and S. Lucey, "Need for speed: A benchmark for higher frame rate object tracking," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1125–1134.

[19] X. Yang, T. Zhang, and C. Xu, "A new discriminative coding method for image classification," *Multimed. Syst.*, pp. 133–145, 2015, doi: 10.1007/s00530-014-0376-y.

[20] J. Zhang, Y. Han, and J. Jiang, "Semi-supervised tensor learning for image classification," *Multimed. Syst.*, pp. 1–11, 2014.

[21] A. Sinha, S. Banerji, and C. Liu, "New color GPHOG descriptors for object and scene image classification," *Mach. Vis. Appl.*, vol. 25, no. 2, pp. 361–375, 2014.

[22] N. Singhai and S. K. Shandilya, "A Survey On:'Content Based Image Retrieval Systems,'" *Int. J. Comput. Appl.*, vol. 4, no. 2, pp. 22–26, 2010.

[23] P. Bolettieri *et al.*, "CoPhIR: a test collection for content-based image retrieval," *ArXiv Prepr. ArXiv09054627*, 2009.

[24] T. Deserno, S. Antani, and L. Rodney Long, "Content-based image retrieval for scientific literature access," *Methods Inf. Med.*, vol. 48, no. 4, p. 371, 2009.

[25] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proceedings of the IEEE International Conference on Computer Vision*, Nov. 2011, pp. 2564–2571, doi: 10.1109/ICCV.2011.6126544.

[26] G. Takacs, V. Chandrasekhar, S. Tsai, D. Chen, R. Grzeszczuk, and B. Girod, "Unified real-time tracking and recognition with rotation-invariant fast features," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 934–941.

[27] D. Debnath and R. Parekh, "Content Based Image Retrieval Using Directional Color Correlograms," *Int. J. Eng. Sci.*, vol. 3, 2011.

[28] C. Taranto, N. Di Mauro, S. Ferilli, and F. Esposito, "Approximate image color correlograms," in *Proceedings of the international conference on Multimedia*, 2010, pp. 1127–1130.

[29] R. Alaoui, S. O. El Alaoui, and M. Meknassi, "Spatial color indexing: An efficient and robust technique for content-based image retrieval," *J. Comput. Sci.*, vol. 5, no. 2, p. 109, 2009.

[30] Y. M. Ro, M. Kim, H. K. Kang, B. Manjunath, and J. Kim, "MPEG-7 homogeneous texture descriptor," *ETRI J.*, vol. 23, no. 2, pp. 41–51, 2001.

[31] K. L. Lee and L. H. Chen, "An efficient computation method for the texture browsing descriptor of MPEG-7," *Image Vis. Comput.*, vol. 23, no. 5, pp. 479–489, 2005.

[32] S. Rahman, S. M. Naim, A. A. Farooq, and M. M. Islam, "Performance of PCA based semi-supervised learning in face recognition using MPEG-7 edge histogram descriptor," *J. Multimed.*, vol. 6, no. 5, pp. 404–415, 2011.

[33] D. Zhang and G. Lu, "Evaluation of MPEG-7 shape descriptors against other shape descriptors," *Multimed. Syst.*, vol. 9, no. 1, pp. 15–30, 2003.

[34] W. Hsu, T.-S. Chua, and H. K. Pung, "An Integrated Color-Spatial Approach to Content-Based Image Retrieval," in *ACM Multimedia*, 1995, pp. 305–313.

[35] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, vol. 1, pp. 886–893.

[36] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *Proceedings of the 6th ACM international conference on Image and video retrieval*, 2007, pp. 401–408.

[37] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, Jan. 2004.

[38] W. Yang *et al.*, "Structure from Motion on XSlit Cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2019, doi: 10.1109/TPAMI.2019.2957119.

[39] H. Li and G. Hua, "Probabilistic Elastic Part Model: A Pose-Invariant Representation for Real-World Face Verification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 918–930, 2018, doi: 10.1109/TPAMI.2017.2695183.

[40] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, Sep. 2008.

[41] Luo Juan and Oubong Gwun, "A Comparison of SIFT, PCA-SIFT and SURF," *Int. J. Image Process.*, vol. 3, no. 4, pp. 143–152, 2009.

[42] X. Yuan, J. Yu, Z. Qin, and T. Wan, "A SIFT-LBP image retrieval model based on bag-of features," in *International Conference on Image Processing (ICIP)*, 2011, pp. 1061–1064.

[43] S. Hwang, "Bag-Of-Visual-Words Approach based on SURF Features to Polyp Detection in Wireless Capsule Endoscopy Videos," Jul. 2011, vol. 2, pp. 941–944.

[44] F. Yang, H. Lu, W. Zhang, and G. Yang, "Visual tracking via bag of features," *Image Process. IET*, vol. 6, no. 2, pp. 115–128, 2012, doi: 10.1049/iet-ipr.2010.0127.

[45] M. S. Martin Stommel, "Binarising SIFT-descriptors to reduce the curse of dimensionality in histogram-based object recognition," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 3, no. 1, pp. 25–36, 2010.

[46] R. Chakraborty, L. Yang, S. Hauberg, and B. Vemuri, "Intrinsic Grassmann Averages for Online Linear, Robust and Nonlinear Subspace Learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2020, doi: 10.1109/TPAMI.2020.2992392.

[47] Y. Xu, G. Feng, and Y. Zhao, "One improvement to two-dimensional locality preserving projection method for use with face recognition," *Neurocomputing*, vol. 73, no. 1, pp. 245–249, 2009.

[48] R. Kapoor and R. Gupta, "Non-linear dimensionality reduction using fuzzy lattices," *Comput. Vis. IET*, vol. 7, no. 3, p., 2013, doi: 10.1049/iet-cvi.2012.0097.

[49] X. Zhou, X. Zhuang, H. Tang, M. Hasegawa-Johnson, and T. S. Huang, "Novel Gaussianized vector representation for improved natural scene categorization," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 702–708, 2010.

[50] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004, vol. 2, pp. 497–506.

[51] L. Akarun, Y. Yardunci, and A. E. Cetin, "Adaptive methods for dithering color images," *Image Process. IEEE Trans. On*, vol. 6, no. 7, pp. 950–955, 1997.

[52] L. Ranathunga, R. Zainuddin, and N. A. Abdullah, "Compacted Dither Pattern Codes Over Mpeg-7 Dominant Colour Descriptor In Video Visual Depiction," *Malays. J. Comput. Sci.*, vol. 23, no. 2, pp. 68–84, 2010.

[53] L. Ranathunga, R. Zainuddin, and N. A. Abdullah, "Performance evaluation of the combination of Compacted Dither Pattern Codes with Bhattacharyya classifier in video visual concept depiction," *Multimed. Tools Appl.*, vol. 54, no. 2, pp. 263–289, 2011.

[54] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[55] M. Paulin, M. Douze, Z. Harchaoui, J. Mairal, F. Perronin, and C. Schmid, "Local convolutional features with unsupervised training for image retrieval," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 91–99.

[56] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *ArXiv Prepr. ArXiv14091556*, 2014.

[57] K. Lin, J. Lu, C.-S. Chen, and J. Zhou, "Learning compact binary descriptors with unsupervised deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1183–1192.

[58] H. Arora, N. Loeff, D. A. Forsyth, and N. Ahuja, "Unsupervised Segmentation of Objects using Efficient Learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2007, pp. 1–7, doi: 10.1109/CVPR.2007.383011.

[59] S. Gould, T. Gao, and D. Koller, "Region-based segmentation and object detection," in *Advances in neural information processing systems*, 2009, pp. 655–663.

[60] A. Vögele, B. Krüger, and R. Klein, "Efficient Unsupervised Temporal Segmentation of Human Motion," Copenhagen, Denmark, Jul. 2014.

[61] K.- Maninis *et al.*, "Video Object Segmentation without Temporal Information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 6, pp. 1515–1530, 2019, doi: 10.1109/TPAMI.2018.2838670.

[62] S. W. Oh, J. Lee, N. Xu, and S. J. Kim, "Fast User-Guided Video Object Segmentation by Interaction-And-Propagation Networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 5242–5251, doi: 10.1109/CVPR.2019.00539.

[63] M. E. Aallaoui and A. Gourch, "Spatio-temporal segmentation with Mumford-Shah functional," in *Computer Systems and Applications (AICCSA), 2013 ACS International Conference on*, 2013, pp. 1–4.

[64] R. Li, S. Yu, and X. Yang, "Efficient spatio-temporal segmentation for extracting moving objects in video sequences," *Consum. Electron. IEEE Trans. On*, vol. 53, no. 3, pp. 1161–1167, 2007.

[65] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[66] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1520–1528.

[67] O. A. Abbas, "Comparisons Between Data Clustering Algorithms.," *Int Arab J Inf Technol*, vol. 5, no. 3, pp. 320–325, 2008.

[68] X. Jin and J. Han, "Expectation Maximization Clustering," in *Encyclopedia of Machine Learning*, Springer, 2010, pp. 382–383.

[69] X. Han, T. Leung, Y. Jia, R. Sukthankar, and A. C. Berg, "Matchnet: Unifying feature and metric learning for patch-based matching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3279–3286.

[70] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4353–4361.

[71] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," in *European conference on computer vision*, 2010, pp. 778–792.

[72] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *2011 International conference on computer vision*, 2011, pp. 2548–2555.

[73] A. Alahi, R. Ortiz, and P. Vandergheynst, "Freak: Fast retina keypoint," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 510–517.

[74] K. Lin, J. Lu, C.-S. Chen, J. Zhou, and M.-T. Sun, "Unsupervised Deep Learning of Compact Binary Descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2018.

[75] C. Strecha, A. Bronstein, M. Bronstein, and P. Fua, "LDAHash: Improved matching with smaller descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 66–78, 2012.

[76] T. Trzcinski and V. Lepetit, "Efficient discriminative projections for compact binary descriptors," in *European Conference on Computer Vision*, 2012, pp. 228–242.

[77] B. Fan, Q. Kong, T. Trzcinski, Z. Wang, C. Pan, and P. Fua, "Receptive fields selection for binary feature description," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2583–2595, 2014.

[78] X. Yang and K.-T. Cheng, "Local difference binary for ultrafast and distinctive feature description," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 188–194, 2014.

[79] T. Trzcinski, M. Christoudias, P. Fua, and V. Lepetit, "Boosting binary keypoint descriptors," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 2874–2881.

[80] A. Andoni and P. Indyk, "Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions," in *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, 2006, pp. 459–468.

[81] R. Salakhutdinov and G. Hinton, "Semantic hashing," *Int. J. Approx. Reason.*, vol. 50, no. 7, pp. 969–978, 2009.

[82] R. Xia, Y. Pan, H. Lai, C. Liu, and S. Yan, "Supervised hashing for image retrieval via image representation learning.," in *AAAI*, 2014, vol. 1, p. 2.

[83] Y. Tian, B. Fan, and F. Wu, "L2-Net: Deep Learning of Discriminative Patch Descriptor in Euclidean Space.," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, vol. 1, pp. 661–669.

[84] H. Liu, R. Wang, S. Shan, and X. Chen, "Deep supervised hashing for fast image retrieval," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2064–2072.

[85] H. Lai, Y. Pan, Y. Liu, and S. Yan, "Simultaneous feature learning and hash coding with deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3270–3278.

[86] M. Lin, Q. Chen, and S. Yan, "Network in network," *ArXiv Prepr. ArXiv13124400*, 2013.

[87] R. Bandara, L. Ranathunga, and N. A. Abdullah, "Invariant Properties of a Locally Salient Dither Pattern with a Spatial-Chromatic Histogram," Peradeniya, Sri Lanka, Dec. 2013.

[88] "Caltech." http://www.vision.caltech.edu/archive.html.

[89] James Z. Wang, "Corel 1000 and Corel 10000 image database." http://wang.ist.psu.edu/docs/related.shtml.

[90] D. Zhang and G. Lu, "A comparative study on shape retrieval using Fourier descriptors with different shape signatures," in *Proceedings of international conference on intelligent multimedia and distance education*, 2001, pp. 1–9.

[91] D. Zhang and G. Lu, "Content-Based Shape Retrieval Using Different Shape Descriptors: A Comparative Study.," in *IEEE International Conference on Multimedia and Expo*, 2001, pp. 1139–1142, doi: 10.1109/ICME.2001.1237928.

[92] S. Shaila and A. Vadivel, "Content-Based Image Retrieval Using Modified Human Colour Perception Histogram," *ITCS SIP JSE-2012 CS IT*, vol. 4, pp. 229–237, 2012.

[93] G. D. Finlayson and G. Schaefer, "Hue that is invariant to brightness and gamma.," in *BMVC*, 2001, pp. 1–10.

[94] V. Ostromoukhov, "A simple and efficient error-diffusion algorithm," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001, pp. 567–572.

[95] A. Bordes, S. Ertekin, J. Weston, and L. Bottou, "Fast kernel classifiers with online and active learning," *J. Mach. Learn. Res.*, vol. 6, no. Sep, pp. 1579–1619, 2005.

[96] G. J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic Object Classes in Video: A High-Definition Ground Truth Database," *Pattern Recognit. Lett.*, vol. 30, no. 2, pp. 88–97, 2008.

[97] A. Papazoglou and V. Ferrari, "Fast object segmentation in unconstrained video," in *Computer Vision (ICCV), 2013 IEEE International Conference on*, 2013, pp. 1777–1784.

[98] A. Jain, S. Chatterjee, and R. Vidal, "Coarse-to-Fine Semantic Video Segmentation Using Supervoxel Trees," in *IEEE International Conference on Computer Vision*, Dec. 2013, pp. 1865–1872, doi: 10.1109/ICCV.2013.234.

[99] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *Pattern Anal. Mach. Intell. IEEE Trans. On*, vol. 34, no. 11, pp. 2274–2282, 2012.

[100] S. N. Tamgade and V. R. Bora, "Motion vector estimation of video image by pyramidal implementation of Lucas Kanade optical flow," in *International Conference on Emerging Trends in Engineering and Technology*, 2009, pp. 914–917.

[101] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal, "Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2009, pp. 1932–1939.

[102] M. Tkalcic, J. F. Tasic, and others, "Colour spaces: perceptual, historical and applicational background," in *The IEEE Region 8 EUROCON 2003. Computer as a Tool.*, 2003, pp. 304–308.

[103] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, 2002.

[104] Y. Duan, J. Lu, Z. Wang, J. Feng, and J. Zhou, "Learning deep binary descriptor with multi-quantization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1183–1192.

[105] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, "Neural codes for image retrieval," in *European conference on computer vision*, 2014, pp. 584–599.

[106] M. Wollborn and R. Mech, "Refined procedure for objective evaluation of video object generation algorithms," *Doc ISOIEC JTC1SC29WG11 M*, vol. 3448, Mar. 1998.

[107] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.

[108] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[109] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Lost in quantization: Improving particular object retrieval in large scale image databases," 2008.

[110] H. Jegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *European conference on computer vision*, 2008, pp. 304–317.

[111] M. Brown, G. Hua, and S. Winder, "Discriminative learning of local image descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 1, pp. 43–57, 2011.

[112] V. Balntas, K. Lenc, A. Vedaldi, and K. Mikolajczyk, "HPatches: A benchmark and evaluation of handcrafted and learned local descriptors," 2017.

[113] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," University of Toronto, 2009.

[114] G. J. Brostow, J. Shotton, J. Fauqueur, and R. Cipolla, "Segmentation and recognition using structure from motion point clouds," in *European conference on computer vision*, Springer, 2008, pp. 44–57.

[115] J. Liu, P. L. Rosin, X. Sun, J. Xiao, and Z. Lian, "Image-driven unsupervised 3D model co-segmentation," *Vis. Comput.*, vol. 35, no. 6–8, pp. 909–920, 2019.

[116] E. Rosten and T. Drummond, "Machine learning for high speed corner detection," in *Euproean Conference on Computer Vision*, 2006, vol. 1, pp. 430–443.

[117] H. Jegou, F. Perronnin, M. Douze, J. Sánchez, P. Perez, and C. Schmid, "Aggregating local image descriptors into compact codes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1704–1716, 2012.

[118] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 806–813.

[119] J. Wan *et al.*, "Deep learning for content-based image retrieval: A comprehensive study," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 157–166.

[120] G. Yunchao, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2916–2929, 2013.

[121] K. Reddy Mopuri and R. Venkatesh Babu, "Object level deep feature pooling for compact image representation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 62–70.

[122] N. Markuš, I. Pandžić, and J. Ahlberg, "Learning Local Descriptors by Optimizing the Keypoint-Correspondence Criterion: Applications to Face Matching, Learning From Unlabeled Videos and 3D-Shape Retrieval," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 279–290, 2018.

[123] Y. Tian, X. Yu, B. Fan, F. Wu, H. Heijnen, and V. Balntas, "SOSNet: Second order similarity regularization for local descriptor learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11016–11025.

[124] Z. Luo *et al.*, "Geodesc: Learning local descriptors by integrating geometry constraints," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 168–183.

[125] M. arten Björkman, N. Bergström, and D. Kragic, "Detecting, segmenting and tracking unknown objects using multi-label MRF inference," *Comput. Vis. Image Underst.*, vol. 118, pp. 111–127, 2014.