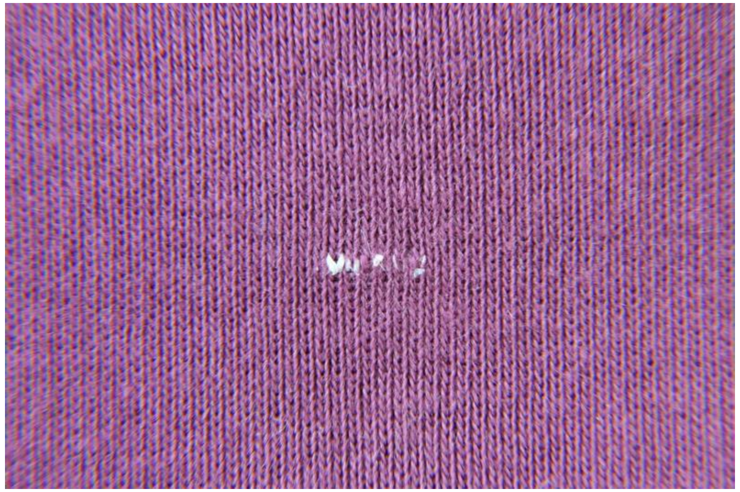
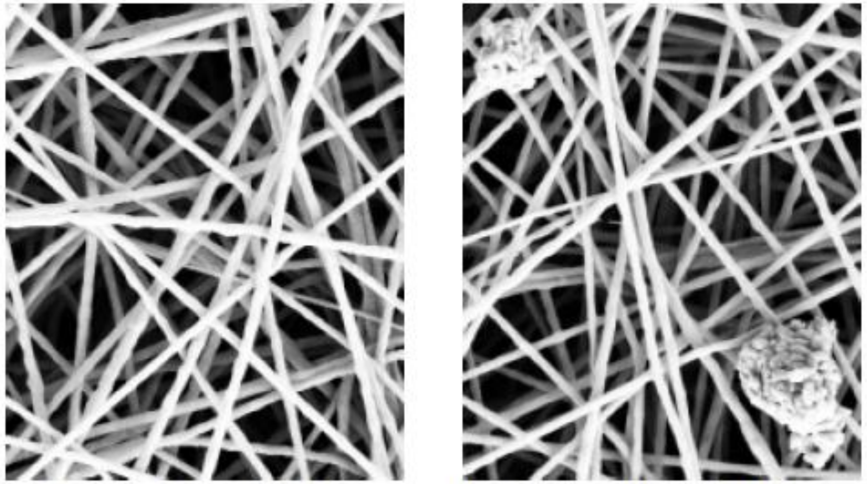
Detection theory and novelty filters

Jean-Michel Morel

In collaboration with
Axel Davy, Tristan Dagobert, Agnès Desolneux, Thibaud Ehret, Rafael Grompone



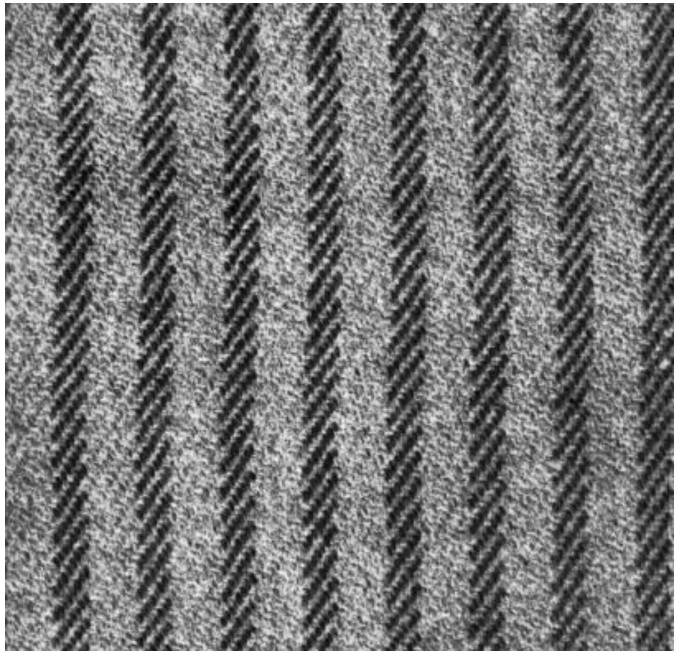
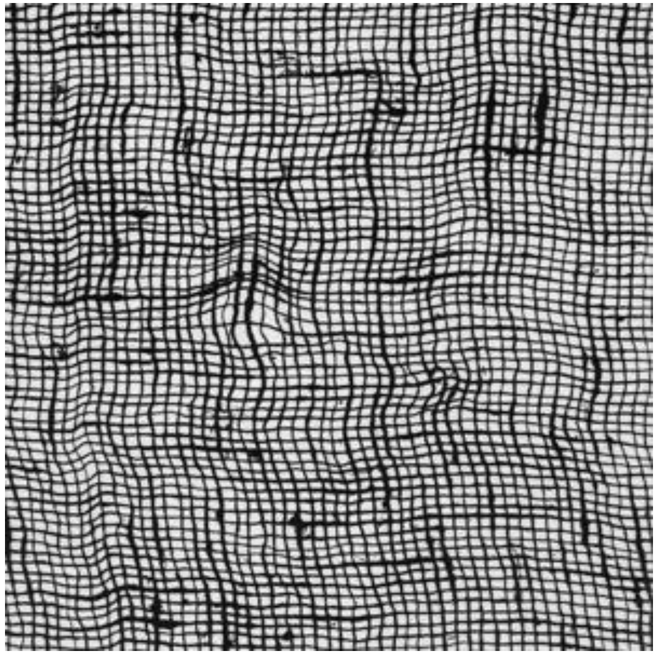
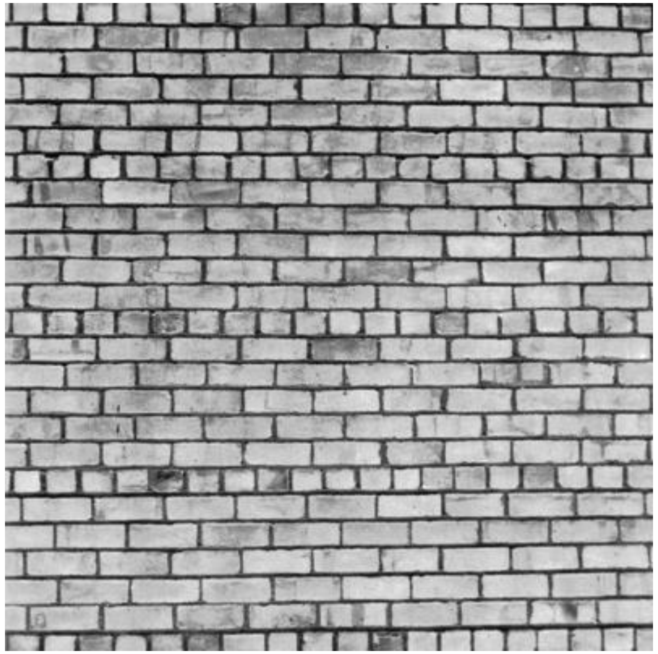


(a) An example of SEM image that is considered normal.

(b) An example of SEM image containing anomalous clots.

Fig. 1. SEM images for monitoring nanofibers production.

Centre de Mathématiques et Leurs Applications
Ecole Normale Supérieure Paris-Saclay (ex de Cachan)

Novelty detection is the task of classifying test data that differ in some respect from the data that are available during “training”.This may be seen as “one-class classification”, in which a model is constructed to describe “normal” data.The novelty detection approach is necessary because the quantity of available “abnormal” data is insufficient to construct explicit models for non-normal classes. In fact novelty detection occurs even in a single image.

This problem then encompasses all methods for estimating a probability density from samples! This is for the « normal » data. The next question is: how far the anomaly is it from being normal? how to decide that it is anomalous?

2. **Probabilistic novelty detection** . . . .

2.1. **Parametric approaches .** . . . .
….2.1.1. Mixture models (e.g. Gaussian Mixtures, estimation by Expectation minimization).. . . . .
….2.1.2. State-space models . . . .

2.2. **Non-parametric** approaches . . . .
…..2.2.1. Kernel density estimators . . . .
…. 2.2.2. Negative selection . . . .

3. Distance-based novelty detection . . . .
…3.1. Nearest neighbor-based approaches . . . .
…3.2. Clustering-based approaches. . . . .

4. Reconstruction-based novelty detection. . . . .
. . . 4.1. Neural network-based approaches . . . .
. . . 4.2. Subspace-based approaches . . . .

A review of novelty detection Marco A.F. Pimentel, David A. Clifton,Lei Clifton, Lionel Tarassenko

Some 1500 references on anomaly/fault/outlier detection…

A review of novelty detection (2014)

Marco A.F.Pimentel, David A.Clifton,Lei Clifton,Lionel Tarassenko

vector machines, in: Proceedings of the International Conference on Machine Learning, 2003, pp. 282–289.

[270] G. Li, C. Wen, Z. Li, A new online learning with kernels method in novelty detection, in: Proceedings of the 37th Annual Conference on IEEE Industrial Electronics Society (IECON), IEEE, 2011, pp. 2311–2316.

[271] L. Manevitz, M. Yousef, One-class SVMs for document classification, J. Mach. Learn. Res. 2 (2002) 139–154.

[272] C. Campbell, K. Bennett, A linear programming approach to novelty detection, in: Proceedings of the Conference on Advances in Neural Information Processing Systems, vol. 13, The MIT Press, 2001, pp. 395–401.

[273] T. Le, D. Tran, W. Ma, D. Sharma, An optimal sphere and two large margins approach for novelty detection, in: Proceedings of the International Joint Conference on Neural Networks (IJCNN), IEEE, 2010, pp. 1–6.

[274] T. Le, D. Tran, W. Ma, D. Sharma, Multiple distribution data description learning algorithm for novelty detection, Adv. Knowl. Discov. Data Min. 6635 (2011) 246–257.

[275] Y.-H. Liu, Y.-C. Liu, Y.-J. Chen, Fast support vector data descriptions for novelty detection, IEEE Trans. Neural Netw. 21 (8) (2010) 1296–1313.

[276] Y.-H. Liu, Y.-C. Liu, Y.-Z. Chen, High-speed inline defect detection for TFT-LCD array process using a novel support vector data description, Exp. Syst. Appl. 38 (5) (2011) 6222–6231.

[277] X. Peng, D. Xu, Efficient support vector data descriptions for novelty detection, Neural Comput. Appl. 21 (8) (2012) 2023–2032.

[278] M. Wu, J. Ye, A small sphere and large margin approach for novelty detection using training data with outliers, IEEE Trans. Pattern Anal. Mach. Intell. 31 (11) (2009) 2088–2092.

32–40.

[293] Z. Wu, W. Xie, J. Yu, Fuzzy c-means clustering algorithm based on kernel method, in: Proceedings of the 5th International Conference on Computational Intelligence and Multimedia Applications (ICCIMA), IEEE, 2003, pp. 49–54.

[294] V. Roth, Outlier detection with one-class kernel fisher discriminants, in: Proceedings of the Conference on Advances in Neural Information Processing Systems (NIPS), 2004.

[295] V. Roth, Kernel fisher discriminants for outlier detection, Neural Comput. 18 (4) (2006) 942–960.

[296] V. Sotiris, P. Tse, M. Pecht, Anomaly detection through a Bayesian support vector machine, IEEE Trans. Reliab. 59 (2) (2010) 277–28

[297] A. Munoz, J. Moguerza, Estimation of high-density regions using one-class neighbor machines, IEEE Trans. Pattern Anal. Mach. Intell. 28 (3) (2006) 476–480.

[298] Y. Li, A surface representation approach for novelty detection, in: Proceedings of the International Conference on Information and Automation (ICIA), IEEE, 2008, pp. 1464–1468.

[299] Z. He, S. Deng, X. Xu, An optimization model for outlier detection in categorical data, Adv. Intell. Comput. 3644 (2005) 400–409.

[300] Z. He, S. Deng, X. Xu, J. Huang, A fast greedy algorithm for outlier mining, Adv. Knowl. Discov. Data Min. 3918 (2006) 567–576.

[301] S. Ando, Clustering needles in a haystack: an information theoretic analysis of minority and outlier detection, in: Proceedings of the 7th IEEE International Conference on Data Mining, ICDM'07, IEEE, 2007, pp. 13–22.

[302] E. Keogh, S. Lonardi, C. Ratanamahatana, Towards parameter-free data mining, in: Proceedings of the 10th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), ACM, 2004, pp.

[258] R. Fujimaki, T. Yairi, K. Machida, An approach to spacecraft anomaly detection problem using kernel feature space, in: Proceedings of the 11th ACM International Conference on Knowledge Discovery in Data Mining (SIGKDD), ACM, 2005, pp. 401–410.

[259] B. Schölkopf, A. Smola, K. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (5) (1998) 1299–1319.

[260] N. Kwak, Principal component analysis based on l1-norm maximization, IEEE Trans. Pattern Anal. Mach. Intell. 30 (9) (2008) 1672–1680.

[261] C. Noble, D. Cook, Graph-based anomaly detection, in: Proceedings of the 9th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), ACM, 2003, pp. 631–636.

[262] J. Sun, H. Qu, D. Chakrabarti, C. Faloutsos, Neighborhood formation and anomaly detection in bipartite graphs, in: Proceedings of the 5th IEEE International Conference on Data Mining, IEEE, 2005, pp. 418–425.

[263] J. Sun, Y. Xie, H. Zhang, C. Faloutsos, Less is more: compact matrix decomposition for large sparse graphs, in: Proceedings of the 7th SIAM International Conference in Data Mining, 2007.

[264] V. Chatzigiannakis, S. Papavassiliou, M. Grammatikou, B. Maglaris, Hierarchical anomaly detection in distributed large-scale sensor networks, in: Proceedings of the 11th IEEE Symposium on Computers and Communications, ISCC'06, IEEE, 2006, pp. 761–

[282] P.F. Evangelista, M.J. Embrechts, B.K. Szymanski, Taming the curse of dimensionality in kernels and novelty detection, in: Applied Soft Computing Technologies: The Challenge of Complexity, Springer Verlag, 2006, pp. 431–444.

[283] A. Gardner, A. Krieger, G. Vachtsevanos, B. Litt, One-class novelty detection for seizure analysis from intracranial EEG, J. Mach. Learn. Res. 7 (2006) 1025–1044.

[284] D.R. Hardoon, L.M. Manevitz, fMRI analysis via one-class machine learning techniques, in: Proceedings of the 19th International Joint Conference on Artificial intelligence, IJCAI'05, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005, pp. 1604–1605.

[285] P. Hayton, S. Utete, D. King, S. King, P. Anuzis, L. Tarassenko, Static and dynamic novelty detection methods for jet engine health monitoring, Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. 365 (1851) (2007) 493–514.

[286] K. Heller, K. Svore, A. Keromytis, S. Stolfo, One class support vector machines for detecting anomalous windows registry accesses, in: Proceedings of the Workshop on Data Mining for Computer Security, 2003.

[287] A. Lazarevic, L. Ertoz, V. Kumar, A. Ozgur, J. Srivastava, A comparative study of anomaly detection schemes in network intrusion detection, in: Proceedings of the 3rd SIAM International Conference on Data Mining, vol. 3, SIAM, 2003, pp. 25–36.

[288] H. Lee, S. Cho, Application of LVQ to novelty detection using outlier

[219] S. Hawkins, H. He, G. Williams, R. Baxter, Outlier detection using replicator neural networks, Data Wareh. Know. Discov. 2454 (2002) 113–123.

[220] N. Japkowicz, Supervised versus unsupervised binary-learning by feedforward neural networks, Mach. Learn. 42 (1) (2001) 97–122.

[221] L. Manevitz, M. Yousef, One-class document classification via neural networks, Neurocomputing 70 (7) (2007) 1466–1481.

[222] B. Thompson, R. Marks, J. Choi, M. El-Sharkawi, M. Huang, C. Bunje, Implicit learning in autoencoder novelty assessment, in: Proceedings of the International Joint Conference on Neural Networks, IJCNN'02, vol. 3, IEEE, 2002, pp. 2878–2883.

[223] G. Williams, R. Baxter, H. He, S. Hawkins, L. Gu, A comparative study of RNN for outlier detection in data mining, in: Proceedings of the IEEE International Conference on Data Mining, IEEE, 2002, pp. 709–712.

[224] S. Jakubek, T. Strasser, Fault-diagnosis using neural networks with ellipsoidal basis functions, in: Proceedings of the American Control Conference, vol. 5, IEEE, 2002, pp. 3846–3851.

[225] Y. Li, M. Pont, N. Barrie Jones, Improving the performance of radial basis function classifiers in condition monitoring and fault diagnosis applications where unknown faults may occur, Pattern Recognit. Lett. 23 (5) (2002) 569–577.

[226] M.K. Albertini, R.F. de Mello, A self-organizing neural network for detecting novelties, in: Proceedings of the 2007 ACM Symposium on Applied Computing, SAC '07, ACM, New York, NY, USA, 2007, pp. 462–466.

[227] G. Barreto, L. Aguayo, Time series clustering for anomaly detection using competitive neural networks, in: J. Principe, R. Miikkulainen (Eds.), Advances in Self-Organizing Maps, Lecture Notes in Computer Science, vol. 5629, Springer, Berlin Heidelberg, 2009, pp. 28–36.

[228] D. Deng, N. Kasabov, On-line pattern analysis by evolving self-organizing maps, Neurocomputing 51 (2003) 87–103.

[229] J. García-Rodríguez, A. Angelopoulou, J. García-Chamizo, A. Psarrou, S. Orts Escolano, V. Morell Giménez, Autonomous growing neural gas for applications with time constraint: optimal parameter estimation, Neural Netw. 32 (2012) 196–208.

[230] D. Hristozov, T. Oprea, J. Gasteiger, Ligand-based virtual screening by novelty detection with self-organizing maps, J. Chem. Inf. Model. 47 (6) (2007) 2044–2062.

analysis and gas sensor arrays under drift conditions, IEEE Sen 6 (3) (2006) 770–783.

[244] T. Ide, H. Kashima, Eigenspace-based anomaly detection in computer systems, in: Proceedings of the 11th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), ACM, 2004, pp. 440–449.

[245] M. Shyu, S. Chen, K. Sarinnapakorn, L. Chang, A Novel Anom Detection Scheme Based on Principal Component Classifier, Te nical Report, DTIC Document, 2003.

[246] M. Thottan, C. Ji, Anomaly detection in IP networks, IEEE Tr Signal Process. 51 (8) (2003) 2191–2204.

[247] J. Toivola, M. Prada, J. Hollmén, Novelty detection in projec spaces for structural health monitoring, Adv. Intell. Data Anal 6065 (2010) 208–219.

[248] Y. Xiao, H. Wang, W. Xu, J. Zhou, L1 norm based KPCA for nov detection, Pattern Recognit. 46 (1) (2013) 389–396.

[249] S. Haggett, D. Chu, I. Marshall, Evolving a dynamic predic coding mechanism for novelty detection, Knowl. Based S 21 (3) (2008) 217–224.

[250] T. Hosoya, S. Baccus, M. Meister, Dynamic predictive coding by retina, Nature 436 (7047) (2005) 71–77.

[251] T. Kohonen, The self-organizing map, Proc. IEEE 78 (9) (19 1464–1480.

[252] K. Labib, R. Vemuri, NSOM: a real-time network-based intrus detection system using self-organizing maps, Netw. Secur. (20 1–6.

[253] D. Alahakoon, S. Halgamuge, B. Srinivasan, Dynamic self-organiz maps with controlled growth for knowledge discovery, IEEE Tr Neural Netw. 11 (3) (2000) 601–614.

[254] J. Blackmore, R. Miikkulainen, Incremental grid growing: encod high-dimensional structure into a two-dimensional feature m in: Proceedings of the IEEE International Conference on Ne Networks, vol. 1, 1993, pp. 450–455.

[255] B. Fritzke, Growing cell structures – a self-organizing network unsupervised and supervised learning, Neural Netw. 7 (9) (19 1441–1460.

[256] B. Fritzke, A growing neural gas network learns topologies, A Neural Inf. Process. Syst. 7 (1995) 625–632.

[257] I. Jolliffe, MyiLibrary, Principal Component Analysis, vol. 2, Wiley Online Library, 2002.

[207] N. Wu, J. Zhang, Factor analysis based anomaly detection, in: Proceedings of the Information Assurance Workshop, IEEE Systems, Man and Cybernetics Society, IEEE, 2003, pp. 108–115.

[208] L. Ertöz, M. Steinbach, V. Kumar, Finding topics in collections of documents: a shared nearest neighbor approach, Clust. Inf. Retr. 11 (2003) 83–103.

[209] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, J. Platt, Support vector method for novelty detection, Adv. Neural Inf. Process. Syst. 12 (3) (2000) 582–588.

[210] J. Zhou, Y. Fu, C. Sun, Y. Fang, Unsupervised distributed novelty detection on scientific simulation data, J. Comput. Inf. Syst. 7 (5) (2011) 1533–1540.

[211] E. Spinosa, A. deLeon, F. de Carvalho, J. Gama, Novelty detection with application to data streams, Intell. Data Anal. 13 (3) (2009) 405–422.

[212] A.F. Hassan, H.M.O. Mokhtar, O. Hegazy, A heuristic approach for sensor network outlier detection, Int. J. Res. Rev. Wirel. Sensor Netw. (IJRRWSN) 1 (4) (2012) 66–72.

[213] T. Idé, S. Papadimitriou, M. Vlachos, Computing correlation anomaly scores using stochastic nearest neighbors, in: Proceedings of the 7th IEEE International Conference on Data Mining (ICDM), IEEE, 2007, pp. 523–528.

[214] K. Onuma, H. Tong, C. Faloutsos, Tangent: a novel,'surprise me', recommendation algorithm, in: Proceedings of the 15th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), ACM, 2009, pp. 657–666.

[215] M. Augusteijn, B. Folkert, Neural network classification and novelty detection, Int. J. Remote Sens. 23 (14) (2002) 2891–2902.

[216] S. Singh, M. Markou, An approach to novelty detection applied to the classification of image regions, IEEE Trans. Knowl. Data Eng. 16 (4) (2004) 396–407.

[217] P. Crook, S. Marsland, G. Hayes, U. Nehmzow, A tale of two filters-on-line novelty detection, in: Proceedings of the IEEE International Conference on Robotics and Automation, ICRA'02, vol. 4, IEEE, 2002, pp. 3894–3899.

[218] I. Diaz, J. Hollmen, Residual generation and visualization for understanding novel process conditions, in: Proceedings of the International Joint Conference on Neural Networks, IJCNN'02, vol. 3, IEEE, 2002, pp. 2070–2075.

[219] S. Hawkins, H. He, G. Williams, R. Baxter, Outlier detection using

[231] D. Kit, B. Sullivan, D. Ballard, Novelty detection using growing neural gas for visuo-spatial memory, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2011, pp. 1194–1200.

[232] S. Marsland, J. Shapiro, U. Nehmzow, A self-organising network that grows when required, Neural Netw. 15 (8–9) (2002) 1041–1058.

[233] S. Marsland, U. Nehmzow, J. Shapiro, On-line novelty detection for autonomous mobile robots, Robot. Auton. Syst. 51 (2) (2005) 191–206.

[234] M. Ramadas, S. Ostermann, B. Tjaden, Detecting anomalous network traffic with self-organizing maps, in: Recent Advances in Intrusion Detection, Springer, 2003, pp. 36–54.

[235] F. Wu, T. Wang, J. Lee, An online adaptive condition-based maintenance method for mechanical systems, Mech. Syst. Signal Process. 24 (8) (2010) 2985–2995.

[236] Y. Chen, B. Malin, Detection of anomalous insiders in collaborative environments via relational analysis of access logs, in: Proceedings of the 1st ACM Conference on Data and Application Security and Privacy, ACM, 2011, pp. 63–74.

[237] Y. Chen, S. Nyemba, B. Malin, Detecting anomalous insiders in collaborative information systems, IEEE Trans. Dependable Secur. Comput. 9 (3) (2012) 332–344.

[238] S. Günter, N. Schraudolph, S. Vishwanathan, Fast iterative kernel principal component analysis, J. Mach. Learn. Res. 8 (2007) 1893–1918.

[239] H. Hoffmann, Kernel PCA for novelty detection, Pattern Recognit. 40 (3) (2007) 863–874.

[240] A. Lakhina, M. Crovella, C. Diot, Mining anomalies using traffic feature distributions, ACM SIGCOMM Comput. Commun. Rev. 35 (4) (2005) 217–228.

[241] R. Kassab, F. Alexandre, Incremental data-driven learning of a novelty detection model for one-class classification with application to high-dimensional noisy data, Mach. Learn. 74 (2) (2009) 191–234.

[242] J. McBain, M. Timusk, Feature extraction for novelty detection as applied to fault detection in machinery, Pattern Recognit. Lett. 32 (7) (2011) 1054–1061.

[243] A. Perera, N. Papamichail, N. Bârsan, U. Weimar, S. Marco, On-line novelty detection by recursive dynamic principal component analysis and gas sensor arrays under drift conditions, IEEE Sens. J.

A review of novelty detection (2014)

Marco A.F. Pimentel, David A. Clifton, Lei Clifton, Lionel Tarassenko

[168] G. Palshikar, Distance-based outliers in sequences, Distrib. Comput. Internet Technol. 3816 (2005) 547–552.

[169] D. Pokrajac, A. Lazarevic, L. Latecki, Incremental local outlier detection for data streams, in: Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM), IEEE, 2007, pp. 504–515.

[170] M. Wu, C. Jermaine, Outlier detection by sampling with accuracy guarantees, in: Proceedings of the 12th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), ACM, 2006, pp. 767–772.

[171] J. Zhang, H. Wang, Detecting outlying subspaces for high-dimensional data: the new task, and performance, Knowl. Inf. Syst. 10 (3) (2006) 333–355.

[172] D. Barbará, Y. Li, J. Couto, COOLCAT: an entropy-based algorithm for categorical clustering, in: Proceedings of the 11th International Conference on Information and Knowledge Management, ACM, 2002, pp. 582–589.

[173] D. Barbará, Y. Li, J. Couto, J. Lin, S. Jajodia, Bootstrapping a data mining intrusion detection system, in: Proceedings of the ACM Symposium on Applied Computing, ACM, 2003, pp. 421–425.

[174] S. Budalakoti, A. Srivastava, R. Akella, E. Turkov, Anomaly Detection in Large Sets of High-Dimensional Symbol Sequences, Technical Report NASA TM-2006-214553, NASA Ames Research Center, 2006.

[175] D. Clifton, P. Bannister, L. Tarassenko, Learning shape for jet engine novelty detection, Adv. Neural Netw. (ISNN) 3973 (2006) 828–835.

[176] D. Clifton, P. Bannister, L. Tarassenko, A framework for novelty detection in jet engine vibration data, Key Eng. Mater. 347 (2007) 305–310.

in: Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, vol. 1215, 1994, pp. 487–499.

[194] S. Papadimitriou, H. Kitagawa, P. Gibbons, C. Faloutsos, LOCI: fast outlier detection using the local correlation integral, in: Proceedings of the 19th International Conference on Data Engineering, IEEE, 2003, pp. 315–326.

[195] A. Chiu, A. Fu, Enhancements on local outlier detection, in: Proceedings of the 7th International Database Engineering and Applications Symposium, IEEE, 2003, pp. 298–307.

[196] J. Tang, Z. Chen, A. Fu, D. Cheung, Enhancing effectiveness of outlier detections for low density patterns, Adv. Knowl. Discov. Data Min. 2336 (2002) 535–548.

[197] D. Ren, B. Wang, W. Perrizo, RDF: a density-based outlier detection method using vertical data representation, in: Proceedings of the 4th IEEE International Conference on Data Mining, ICDM'04, IEEE, 2004, pp. 503–506.

[198] J. Yu, W. Qian, H. Lu, A. Zhou, Finding centric local outliers in categorical/numerical spaces, Knowl. Inf. Syst. 9 (3) (2006) 309–338.

[199] J. Tang, Z. Chen, A. Fu, D. Cheung, Capabilities of outlier detection in large datasets, framework and methodologies, Knowl. Inf. Syst. 11 (1) (2007) 45–84.

[200] P. Sun, S. Chawla, On local spatial outliers, in: Proceedings of the 4th IEEE International Conference on Data Mining, IEEE, 2004, pp. 209–216.

[201] P. Sun, S. Chawla, B. Arunasalam, Mining for outliers in sequential databases, in: Proceedings of the 6th SIAM International Conference on Data Mining, vol. 124, Society for Industrial Mathematics, 2006.

Principles of Data Mining and Knowledge Discovery, PKDD '02, Springer-Verlag, London, UK, 2002, pp. 15–26.

[157] S. Bay, M. Schwabacher, Mining distance-based outliers in near linear time with randomization and a simple pruning rule, in: Proceedings of the 9th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), ACM, 2003, pp. 29–38.

[158] S. Boriah, V. Chandola, V. Kumar, Similarity measures for categorical data: a comparative evaluation, in: Proceedings of the 8th SIAM International Conference on Data Mining, 2008, pp. 243–254.

[159] M. Breunig, H. Kriegel, R. Ng, J. Sander, LOF: identifying density-based local outliers, in: Proceedings of the ACM SIGMOD International Conference on Management of Data, vol. 29, ACM, 2000, pp. 93–104.

[160] V. Chandola, S. Boriah, V. Kumar, Understanding Categorical Similarity Measures for Outlier Detection, Technical Report 08-008, University of Minnesota, 2008.

[161] S. Chawla, P. Sun, SLOM: a new measure for local spatial outliers, Knowl. Inf. Syst. 9 (4) (2006) 412–429.

[162] A. Ghoting, M. Otey, S. Parthasarathy, Loaded: link-based outlier and anomaly detection in evolving data sets, in: Proceedings of the 4th IEEE International Conference on Data Mining, ICDM'04, IEEE, 2004, pp. 387–390.

[163] A. Ghoting, S. Parthasarathy, M. Otey, Fast mining of distance-based outliers in high-dimensional datasets, Data Min. Knowl. Discov. 16 (3) (2008) 349–364.

[164] V. Hautamaki, I. Karkkainen, P. Franti, Outlier detection using k-nearest neighbour graph, in: Proceedings of the 17th Interna-

techniques, in: Proceedings of the IEEE Aerospace Conference, IEEE, 2006, pp. 1–17.

[182] H. Sun, Y. Bao, F. Zhao, G. Yu, D. Wang, CD-trees: an efficient index structure for outlier detection, Adv. Web-Age Inf. Manage. 3129 (2004) 600–609.

[183] Z. Syed, M. Saeed, I. Rubinfeld, Identifying high-risk patients without labeled training data: anomaly detection methodologies to predict adverse outcomes, in: AMIA Annual Symposium Proceedings, vol. 2010, American Medical Informatics Association, 2010, pp. 772–776.

[184] C.-H. Wang, Outlier identification and market segmentation using kernel-based clustering techniques, Exp. Syst. Appl. 36 (2) (2009) 3744–3750.

[185] J. Yang, W. Wang, CLUSEQ: efficient and effective sequence clustering, in: Proceedings of the 19th International Conference on Data Engineering, IEEE, 2003, pp. 101–112.

[186] S.-P. Yong, J.D. Deng, M.K. Purvis, Novelty detection in wildlife scenes through semantic context modelling, Pattern Recognit. 45 (9) (2012) 3439–3450.

[187] S. Yong, J. Deng, M. Purvis, Wildlife video key-frame extraction based on novelty detection in semantic context, Multimed. Tools Appl. 62 (2) (2013) 359–376.

[188] D. Yu, G. Sheikholeslami, A. Zhang, Findout: finding outliers in very large datasets, Knowl. Inf. Syst. 4 (4) (2002) 387–412.

[189] K. Zhang, S. Shi, H. Gao, J. Li, Unsupervised outlier detection in sensor networks using aggregation tree, Adv. Data Min. Appl. 4632 (2007) 158–169.

on a laboratory structure, J. Sound Vib. 259 (2) (2003) 323–343.

[117] K. Yamanishi, J. Takeuchi, G. Williams, P. Milne, On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms, Data Min. Knowl. Discov. 8 (3) (2004) 275–300.

[118] K. Yamanishi, J. Takeuchi, G. Williams, P. Milne, On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms, in: Proceedings of the 6th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), ACM, 2000, pp. 320–324.

[119] D. Agarwal, An empirical Bayes approach to detect anomalies in dynamic multidimensional arrays, in: Proceedings of the 5th IEEE International Conference on Data Mining, IEEE, 2005, pp. 26–33.

[120] D. Agarwal, Detecting anomalies in cross-classified streams: a Bayesian approach, Knowl. Inf. Syst. 11 (1) (2007) 29–44.

[121] F. Zorriassatine, J. Tannock, C. O'Brien, Using novelty detection to identify abnormalities caused by mean shifts in bivariate processes, Comput. Ind. Eng. 44 (3) (2003) 385–408.

[122] P. Højen-Sørensen, O. Winther, L. Hansen, Mean-field approaches to independent component analysis, Neural Comput. 14 (4) (2002) 889–918.

[123] J. Verbeek, N. Vlassis, B. Kröse, Efficient greedy learning of Gaussian mixture models, Neural Comput. 15 (2) (2003) 469–485.

[124] J. Zhang, Z. Ghahramani, Y. Yang, A probabilistic model for online document clustering with application to novelty detection, in: NIPS, 2005.

[125] P. Perner, Concepts for novelty detection and handling based on a case-based reasoning process scheme, Eng. Appl. Artif. Intell. 22 (1) (2009) 86–91.

[126] K. Hempstalk, E. Frank, I. Witten, One-class classification by combining density and class probability estimation, in: W. Daelemans, B. Goethals, K. Morik (Eds.), Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science, vol. 5211, Springer, Berlin/Heidelberg, 2008, pp. 505–519.

[127] D. Chen, M. Meng, Health status detection for patients in physiological monitoring, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2011, pp. 4921–4924.

[128] T. Kanamori, S. Hido, M. Sugiyama, A least-squares approach to direct importance estimation, J. Mach. Learn. Res. 10 (2009) 1391–1445.

network traffic for detecting novel attacks, in: Proceedings of the 8th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), ACM, 2002, pp. 376–385.

[143] E. Parzen, On estimation of a probability density function and mode, Ann. Math. Stat. 33 (3) (1962) 1065–1076.

[144] A. Frank, A. Asuncion, UCI machine learning repository, 2010.

[145] M. Hravnak, L. Edwards, A. Clontz, C. Valenta, M. DeVita, M. Pinsky, Defining the incidence of cardiorespiratory instability in patients in step-down units using an electronic integrated monitoring system, Arch. Internal Med. 168 (12) (2008) 1300–1308.

[146] P. Angelov, D. Filev, An approach to online identification of Takagi-Sugeno fuzzy models, IEEE Trans. Syst. Man Cybern. Part B: Cybern. 34 (1) (2004) 484–498.

[147] R. Adams, I. Murray, D. MacKay, The Gaussian process density sampler, in: Advances in Neural Information Processing Systems (NIPS) 21, 2009, pp. 9–16.

[148] G. Lorden, Procedures for reacting to a change in distribution, Ann. Math. Stat. 42 (6) (1971) 1897–1908.

[149] M. Basseville, I.V. Nikiforov, Detection of Abrupt Changes: Theory and Application, vol. 104, Prentice Hall, Englewood Cliffs, 1993.

[150] J. Reeves, J. Chen, X.L. Wang, R. Lund, Q.Q. Lu, A review and comparison of changepoint detection techniques for climate data, J. Appl. Meteorol. Climatol. 46 (6) (2007) 900–915.

[151] T. Peng, C. Leckie, K. Ramamohanarao, Information sharing for distributed intrusion detection systems, J. Netw. Comput. Appl. 30 (3) (2007) 877–899.

[152] T. Van Phuong, L. Hung, S. Cho, Y. Lee, S. Lee, An anomaly detection algorithm for detecting attacks in wireless sensor networks, Intell. Secur. Informat. 3975 (2006) 735–736.

[153] A.G. Tartakovsky, G.V. Moustakides, State-of-the-art in Bayesian changepoint detection, Seq. Anal. 29 (2) (2010) 125–145.

[154] J. Chen, A.K. Gupta, Parametric Statistical Change Point Analysis: with Applications to Genetics, Birkhäuser Boston, Medicine, and Finance, Boston, 2012.

[155] S. Forrest, A. Perelson, L. Allen, R. Cherukuri, Self-nonself discrimination in a computer, in: Proceedings of the IEEE Computer Society Symposium on Research in Security and Privacy, IEEE, 1994, pp. 202–212.

[156] F. Angiulli, C. Pizzuti, Fast outlier detection in high dimensional spaces, in: Proceedings of the 6th European Conference on

Proceedings of the 3rd IASTED Conference on Artificial Intelligence and Applications, 2003, pp. 457–462.

[105] I. Mayrose, N. Friedman, T. Pupko, A gamma mixture model better accounts for among site rate heterogeneity, Bioinformatics 21 (2) (2005) 151–158.

[106] A. Carvalho, M. Tanner, Modelling nonlinear count time series with local mixtures of poisson autoregressions, Comput. Stat. Data Anal. 51 (11) (2007) 5266–5294.

[107] M. Svensén, C. Bishop, Robust Bayesian mixture modelling, Neurocomputing 64 (2005) 235–252.

[108] A. Stranjak, P. Dutta, M. Ebden, A. Rogers, P. Vytelingum, A multi-agent simulation system for prediction and scheduling of aero engine overhaul, in: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems: Industrial Track, International Foundation for Autonomous Agents and Multiagent Systems, 2008, pp. 81–88.

[109] L. Parra, G. Deco, S. Miesbach, Statistical independence and novelty detection with information preserving nonlinear maps, Neural Comput. 8 (2) (1996) 260–269.

[110] A. Nairac, T. Corbett-Clark, R. Ripley, N. Townsend, L. Tarassenko, Choosing an appropriate model for novelty detection, in: Proceedings of the 5th International Conference on Artificial Neural Networks, IET, 1997, pp. 117–122.

[111] R. Hyndman, Computing and graphing highest density regions, Am. Stat. 50 (2) (1996) 120–126.

[112] J. Pickands, Statistical inference using extreme order statistics, Ann. Stat. 3 (1) (1975) 119–131.

[113] P. Embrechts, C. Klüppelberg, T. Mikosch, Modelling Extremal Events for Insurance and Finance, vol. 33, Springer Verlag, 1997.

[114] R. Fisher, L. Tippett, Limiting forms of the frequency distribution of the largest or smallest member of a sample, in: Proceedings of the Cambridge Philosophical Society, vol. 24, Cambridge University Press, 1928, pp. 180–190.

[115] D. Clifton, L. Tarassenko, N. McGrogan, D. King, S. King, P. Anuzis, Bayesian extreme value statistics for novelty detection in gas-turbine engines, in: Proceedings of the IEEE Aerospace Conference, IEEE, 2008, pp. 1–11.

[116] K. Worden, G. Manson, D. Allman, Experimental validation of a structural health monitoring methodology: part i. Novelty detection on a laboratory structure, J. Sound Vib. 259 (2) (2003) 323–343.

[129] S. Hido, Y. Tsuboi, H. Kashima, M. Sugiyama, T. Kanamori, Statistical outlier detection using direct density ratio estimation, Knowl. Inf. Syst. 26 (2) (2011) 309–336.

[130] M. Sugiyama, T. Suzuki, T. Kanamori, Density ratio estimation: a comprehensive review, RIMS Kokyuroku (2010) 10–31.

[131] S. Hoare, D. Asbridge, P. Beatty, On-line novelty detection for artefact identification in automatic anaesthesia record keeping, Med. Eng. Phys. 24 (10) (2002) 673–681.

[132] S. Roberts, L. Tarassenko, A probabilistic resource allocating network for novelty detection, Neural Comput. 6 (2) (1994) 270–284.

[133] P. Galeano, D. Peña, R. Tsay, Outlier detection in multivariate time series by projection pursuit, J. Am. Stat. Assoc. 101 (474) (2006) 654–669.

[134] D. Chen, X. Shao, B. Hu, Q. Su, Simultaneous wavelength selection and outlier detection in multivariate regression of near-infrared spectra, Anal. Sci. 21 (2) (2005) 161–166.

[135] K. Kadota, D. Tominaga, Y. Akiyama, K. Takahashi, Detecting outlying samples in microarray data: a critical assessment of the effect of outliers on sample classification, Chem-Bio Informat. 3 (1) (2003) 30–45.

[136] P. Smyth, Markov monitoring with unknown states, IEEE J. Sel. Areas Commun. 12 (9) (1994) 1600–1612.

[137] Z. Ghahramani, G. Hinton, Variational learning for switching state-space models, Neural Comput. 12 (4) (2000) 831–864.

[138] M. Atallah, W. Szpankowski, R. Gwadera, Detection of significant sets of episodes in event sequences, in: Proceedings of the 4th IEEE International Conference on Data Mining, ICDM'04, IEEE, 2004, pp. 3–10.

[139] A. Sebyala, T. Olukemi, L. Sacks, Active platform security through intrusion detection using naive Bayesian network for anomaly detection, in: London Communications Symposium, Citeseer, 2002.

[140] C. Kruegel, G. Vigna, Anomaly detection of web-based attacks, in: Proceedings of the 10th ACM Conference on Computer and Communications Security, ACM, 2003, pp. 251–261.

[141] C. Kruegel, D. Mutz, W. Robertson, F. Valeur, Bayesian event classification for intrusion detection, in: Proceedings of the 19th Annual Computer Security Applications Conference, IEEE, 2003, pp. 14–23.

[142] M. Mahoney, P. Chan, Learning nonstationary models of normal network traffic for detecting novel attacks, in: Proceedings of the

A review of novelty detection (2014)

Marco A.F.Pimentel, David A.Clifton,Lei Clifton,Lionel Tarassenko

[68] R. Gwadera, M. Atallah, W. Szpankowski, Markov models for identification of significant episodes, in: Proceedings of 5th SIAM International Conference on Data Mining, 2005, pp. 404–414.
[69] R. Gwadera, M. Atallah, W. Szpankowski, Reliable detection of episodes in event sequences, Knowl. Inf. Syst. 7 (4) (2005) 415–437.
[70] A. Ihler, J. Hutchins, P. Smyth, Adaptive event detection with time-varying poisson processes, in: Proceedings of the 12th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), ACM, 2006, pp. 207–216.
[71] D. Janakiram, V. Adi Mallikarjuna Reddy, A. Phani Kumar, Outlier detection in wireless sensor networks using Bayesian belief networks, in: Proceedings of the 1st International Conference on Communication System Software and Middleware (Comsware), IEEE, 2006, pp. 1–6.
[72] H.-J. Lee, S. Roberts, On-line novelty detection using the Kalman filter and extreme value theory, in: Proceedings of the 19th International Conference on Pattern Recognition (ICPR), 2008, pp. 1–4.
[73] P. McSharry, T. He, L. Smith, L. Tarassenko, Linear and non-linear methods for automatic seizure detection in scalp electro-

detection in video streams using recursive de
Proceedings of the 4th International IEEE Co
Systems, IS'08, IEEE, vol. 2, 2008, pp. 14–22
[93] S. Subramaniam, T. Palpanas, D. Papadop
D. Gunopulos, Online outlier detection in sen
parametric models, in: Proceedings of the
Conference on Very Large Databases, VLDB
pp. 187–198.
[94] L. Tarassenko, A. Hann, A. Patterson, E. Braith
V. Barber, D. Young, Biosign™: multi-param
early warning of patient deterioration, in: Pro
IEE International Seminar on Medical Applica
cessing, IET, 2005, pp. 71–76.
[95] L. Tarassenko, A. Hann, D. Young, Integrat
analysis for early warning of patient deteriora
97 (1) (2006) 64–68.
[96] P. Vincent, Y. Bengio, Manifold parzen windo
Process. Syst. 15 (2002) 825–832.
[97] D. Yeung, C. Chow, Parzen-window network
in: Proceedings of the 16th International Co
Recognition, vol. 4, IEEE, 2002, pp. 385–388.
[98] D. Dasgupta, N. Majumdar, Anomaly detection i

[58] D. Clifton, S. Hugueny, L. Tarassenko, A comparison of approaches to multivariate extreme value theory for novelty detection, in: Proceedings of the IEEE/SP 15th Workshop on Statistical Signal Processing, IEEE, 2009, pp. 13–16.
[59] D. Clifton, S. Hugueny, L. Tarassenko, Novelty detection with multivariate extreme value statistics, J. Signal Process. Syst. 65 (3) (2011) 371–389.
[60] D. Clifton, S. Hugueny, L. Tarassenko, Pinning the tail on the distribution: a multivariate extension to the generalised Pareto distribution, in: IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2011, pp. 1–6.
[61] D. Clifton, L. Clifton, S. Hugueny, D. Wong, L. Tarassenko, An extreme function theory for novelty detection, IEEE J. Sel. Top. Signal Process. 7 (1) (2013) 28–37.
[62] A. Hazan, J. Lacaille, K. Madani, Extreme value statistics for vibration spectra outlier detection, in: Proceedings of the 9th International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, 2012.
[63] S. Hugueny, D. Clifton, L. Tarassenko, Novelty detection with multivariate extreme value theory, part II: an analytical approach to unimodal estimation, in: Proceedings of the IEEE International

[80] C. Williams, J. Quinn, N. McIntosh, Factorial switching Kalman filters for condition monitoring in neonatal intensive care, Neural Inf. Process. (2006) 1513–1520.
[81] W. Wong, A. Moore, G. Cooper, M. Wagner, Rule-based anomaly pattern detection for detecting disease outbreaks, in: Proceedings of the National Conference on Artificial Intelligence, Menlo Park, CA; Cambridge, MA; London, AAAI Press; MIT Press; 1999, 2002, pp. 217–223.
[82] W. Wong, A. Moore, G. Cooper, M. Wagner, Bayesian network anomaly pattern detection for disease outbreaks, in: Proceedings of the 20th International Conference on Machine Learning, vol. 20, AAAI Press, 2003, pp. 808–815.
[83] D.-Y. Yeung, Y. Ding, Host-based intrusion detection using dynamic and static behavioral models, Pattern Recognit. 36 (1) (2003) 229–243.
[84] X. Zhang, P. Fan, Z. Zhu, A new anomaly detection method based on hierarchical HMM, in: Proceedings of the 4th International Conference on Parallel and Distributed Computing, Applications and Technologies, IEEE, 2003, pp. 249–252.
[85] P. Angelov, An approach for fuzzy rule-base adaptation using on-line clustering, Int. J. Approx. Reason. 35 (3) (2004) 275–289.

Computer Science and Information Systems (FedCSIS), IEEE, 2011, pp. 125–131.
[4] L. Tarassenko, D. Clifton, P. Bannister, S. King, D. King, Novelty Detection, John Wiley & Sons, Ltd, 2009, pp. 1–22 (Chapter 35).
[5] C. Surace, K. Worden, Novelty detection in a changing environment: a negative selection approach, Mech. Syst. Signal Process. 24 (4) (2010) 1114–1128.
[6] A. Patcha, J. Park, An overview of anomaly detection techniques: existing solutions and latest technological trends, Comput. Netw. 51 (12) (2007) 3448–3470.
[7] V. Jyothsna, V.V.R. Prasad, K.M. Prasad, A review of anomaly based intrusion detection systems, Int. J. Comput. Appl. 28 (7) (2011) 26–35.
[8] C. Diehl, J. Hampshire, Real-time object classification and novelty detection for collaborative video surveillance, in: Proceedings of the International Joint Conference on Neural Networks, IJCNN'02, 2002, vol. 3, pp. 2620–2625.
[9] M. Markou, S. Singh, A neural network-based novelty detector for image sequence analysis, IEEE Trans. Pattern Anal. Mach. Intell. 28 (10) (2006) 1664–1677.
[10] H. Vieira Neto, U. Nehmzow, Real-time automated visual inspection using mobile robots, J. Intell. Robotic Syst. 49 (3) (2007) 293–307.
[11] B. Sofman, B. Neuman, A. Stentz, J. Bagnell, Anytime online novelty and change detection for mobile robots, J. Field Robot. 28 (4) (2011) 589–618.
[12] Y. Zhang, N. Meratnia, P. Havinga, Outlier detection techniques for wireless sensor networks: a survey, IEEE Commun. Surv. Tutor. 12 (2) (2010) 159–170.
[13] H. Dutta, C. Giannella, K. Borne, H. Kargupta, Distributed top-k outlier detection from astronomy catalogs using the DEMAC system, in: Proceedings of the 7th SIAM International Conference on Data Mining, IEEE, 2007.
[14] H. Escalante, A comparison of outlier detection algorithms for machine learning, in: Proceedings of the International Conference on Communications in Computing, Citeseer, 2005.
[15] S. Basu, M. Bilenko, R. Mooney, A probabilistic framework for semi-supervised clustering, in: Proceedings of the 10th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), ACM, 2004, pp. 59–68.

[31] Z. Bakar, R. Mohemad, A. Ahmad, M. Deris, A comp
outlier detection techniques in data mining, in: P
IEEE Conference on Cybernetics and Intelligen
2006, pp. 1–6.
[32] S. Khan, M. Madden, A survey of recent tre
classification, in: L. Coyle, J. Freyne (Eds.), Arti
and Cognitive Science, Lecture Notes in Compu
6206, Springer, Berlin/Heidelberg, 2010, pp. 188–
[33] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classific
USA, 2001.
[34] C. Bishop, Pattern Recognition and Machine
Springer, New York, 2006.
[35] A. Modenesi, A. Braga, Analysis of time series r
strategies for synthetic and real data, Neural Pro
(2009) 1–17.
[36] V. Chandola, A. Banerjee, V. Kumar, Outlier Dete
Technical Report 07-017, University of Minnesota
[37] J. Kittler, W. Christmas, T. de Campos, D. Wi
J. Illingworth, M. Osman, Domain anomaly dete
perception: a system architecture and taxono
Pattern Anal. Mach. Intell. 99 (2013) 1.
[38] A.M. Bartkowiak, Anomaly, novelty, one-class
comprehensive introduction, Int. J. Comput. Inf. S
Appl. 3 (2011) 61–71.
[39] Y. Gatsoulis, E. Kerr, J. Condell, N. Siddique, T. M
detection for cumulative learning, in: Proceedin
ence on Towards Autonomous Robotic Systems, 2
[40] E. Kerr, Y. Gatsoulis, N.H. Siddique, J.V. Condell,
Brief overview of novelty detection methods for r
learning, in: Proceedings of the 21st National
Artificial Intelligence and Cognitive Science, 2010
[41] D. Miljkovic, Review of novelty detection methods,
the 33rd International Convention (MIPRO), IEEE, 20
[42] R. Helali, Data mining based network intrusion de
survey, Novel Algoritm. Tech. Telecommun. Netw.
[43] F.E. Grubbs, Procedures for detecting outlying
samples, Technometrics 11 (1) (1969) 1–21.
[44] C. Aggarwal, P. Yu, Outlier detection with un
Proceedings of the SIAM International Conference

[16] D. Barber, Bayesian Reasoning and Machine Learning, Cambridge University Press, 2012.
[17] C. Sammut, G. Webb, Encyclopedia of Machine Learning, Springer, 2011, Springer reference.
[18] M. Moya, M. Koch, L. Hostetler, One-class classifier networks for target recognition applications, in: Proceedings of the World Congress on Neural Networks, International Neural Network Society, 1993, pp. 797–801.
[19] H. He, E. Garcia, Learning from imbalanced data, IEEE Trans. Knowl. Data Eng. 21 (9) (2009) 1263–1284.
[20] H.-J. Lee, S. Cho, The novelty detection approach for different degrees of class imbalance, in: I. King, J. Wang, L.-W. Chan, D. Wang (Eds.), Neural Information Processing, Lecture Notes in Computer Science, vol. 4233, Springer, Berlin/Heidelberg, 2006, pp. 21–30.
[21] C. Bishop, Novelty detection and neural network validation, in: Proceedings of the IEEE Conference on Vision, Image and Signal Processing, vol. 141, IET, 1994, pp. 217–222.
[22] G. Ritter, M. Gallegos, Outliers in statistical pattern recognition and an application to automatic chromosome classification, Pattern Recognit. Lett. 18 (6) (1997) 525–539.
[23] I. Merriam-Webster, Merriam-webster – an encyclopedia britannica company, May 2012. URL ⟨http://www.merriam-webster.com/dictionary/novel/⟩.
[24] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: a survey, ACM Comput. Surv. (CSUR) 41 (3) (2009) 1–58.
[25] V. Barnett, T. Lewis, Outliers in Statistical Data, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, Wiley and Sons, 1994.
[26] M. Markou, S. Singh, Novelty detection: a review – part 1: statistical approaches, Signal Process. 83 (12) (2003) 2481–2497.
[27] M. Markou, S. Singh, Novelty detection: a review – part 2: neural network based approaches, Signal Process. 83 (12) (2003) 2499–2521.
[28] S. Marsland, Novelty detection in learning systems, Neural Comput. Surv. 3 (2003) 157–195.
[29] V. Hodge, J. Austin, A survey of outlier detection methodologies, Artif. Intell. Rev. 22 (2) (2004) 85–126.
[30] M. Agyemang, K. Barker, R. Alhajj, A comprehensive survey of numeric and symbolic outlier mining techniques, Intell. Data Anal. 10 (6) (2006) 521–538.

2008, pp. 483–493.
[45] H. Solberg, A. Lahti, Detection of outliers in reference distributions: performance of Horn's algorithm, Clin. Chem. 51 (12) (2005) 2326–2332.
[46] C. Chow, On optimum recognition error and reject tradeoff, IEEE Trans. Inf. Theory 16 (1) (1970) 41–46.
[47] D.W. Scott, Frontmatter, John Wiley & Sons, Inc., 2008.
[48] D. Filev, F. Tseng, Real time novelty detection modeling for machine health prognostics, in: Proceedings of the Annual Meeting of the North American Fuzzy Information Processing Society (NAFIPS), IEEE, 2006, pp. 529–534.
[49] D. Filev, F. Tseng, Novelty detection based machine health prognostics, in: International Symposium on Evolving Fuzzy Systems, 2006, pp. 193–199.
[50] A. Flexer, E. Pampalk, G. Widmer, Novelty detection based on spectral similarity of songs, in: Proceedings of 6th International Conference on Music Information Retrieval, 2005, pp. 260–263.
[51] J. Ilonen, P. Paalanen, J. Kamarainen, H. Kalviainen, Gaussian mixture pdf in one-class classification: computing and utilizing confidence values, in: Proceedings of the 18th International Conference on Pattern Recognition (ICPR), vol. 2, IEEE, 2006, pp. 577–580.
[52] J. Larsen, Distribution of the Density of a Gaussian Mixture, Technical Report, Informatics and Mathematical Modelling, DTU, 2003.
[53] P. Paalanen, J. Kamarainen, J. Ilonen, H. Kälviäinen, Feature representation and discrimination based on Gaussian mixture model probability densities – practices and algorithms, Pattern Recognit. 39 (7) (2006) 1346–1358.
[54] N. Pontoppidan, J. Larsen, Unsupervised condition change detection in large diesel engines, in: Proceedings of the IEEE 13th Workshop on Neural Networks for Signal Processing, NNSP'03, IEEE, 2003, pp. 565–574.
[55] X. Song, M. Wu, C. Jermaine, S. Ranka, Conditional anomaly detection, IEEE Trans. Knowl. Data Eng. 19 (5) (2007) 631–645.
[56] F. Zorriassatine, A. Al-Habaibeh, R. Parkin, M. Jackson, J. Coy, Novelty detection for practical pattern recognition in condition monitoring of multivariate processes: a case study, Int. J. Adv. Manuf. Technol. 25 (9) (2005) 954–963.
[57] D. Clifton, L. Clifton, P. Bannister, L. Tarassenko, Automated novelty detection in industrial systems, Adv. Comput. Intell. Ind. Syst. 116 (2008) 269–296.

Detection principles illustrated by a few classic methods

MULTISCALE ANOMALY DETECTION USING DIFFUSION MAPS AND SALIENCY SCORE
Gal Mishne and Israel Cohen

Principles 1 and 2: dimension reduction and multiscale detection

A multiscale approach to anomaly detection in images, combining diffusion maps for dimensionality reduction and a nearest-neighbor-based anomaly score in the reduced dimension.

Let $\Gamma = \{x_1, .., x_n\}$ be a high-dimensional set of n data points. A weighted graph is constructed with the data points as nodes and the weights of the edges connecting two node is a measure of the similarity between the two data points. The affinity matrix $\mathbf{W} = w(x_i, x_j)$, $x_i, x_j \in \Gamma$ is required to be symmetric and non-negative. A common choice is an RBF kernel $w(x_i, x_j) = \exp\left\{-\|x_i - x_j\|^2/\sigma^2\right\}$, where $\sigma > 0$ is a scale parameter. Then, a random walk is created on the data set by normalizing the kernel:

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}, \tag{1}$$

where $\mathbf{D}(i,i) = \sum_{j\in\Gamma} w(x_i, x_j)$. The row-stochastic matrix $\mathbf{P}$ satisfies $p(x_i, x_j) \geq 0$ and $\sum_{j\in\Gamma} p(x_i, x_j) = 1$ and can be viewed as the transition matrix of a Markov chain on the data set Γ. The

Euclidean distance in this new embedding. Retaining only the first ℓ eigenvectors, the diffusion map is defined by

$$\Psi_t : x_i \rightarrow \left(\lambda_1^t \psi_1(x_i), \lambda_2^t \psi_2(x_i), ..., \lambda_\ell^t \psi_\ell(x_i)\right)^T. \tag{5}$$

where ϕ_l and ψ_l are the biorthogonal left and right eigenvectors, respectively, and $|\lambda_0| \geq |\lambda_1| \geq ... \geq 0$ are the sequence of eigenvalues.

anomaly score is given by

$$S(i)_{DM} = 1 - \exp\left\{-\frac{1}{K}\sum_{k=1}^{K} \frac{d_{\mathrm{DM}}(p_i, p_j)/2\sigma_K}{1 + c \cdot d_{\mathrm{position}}(p_i, p_j)}\right\}. \tag{9}$$

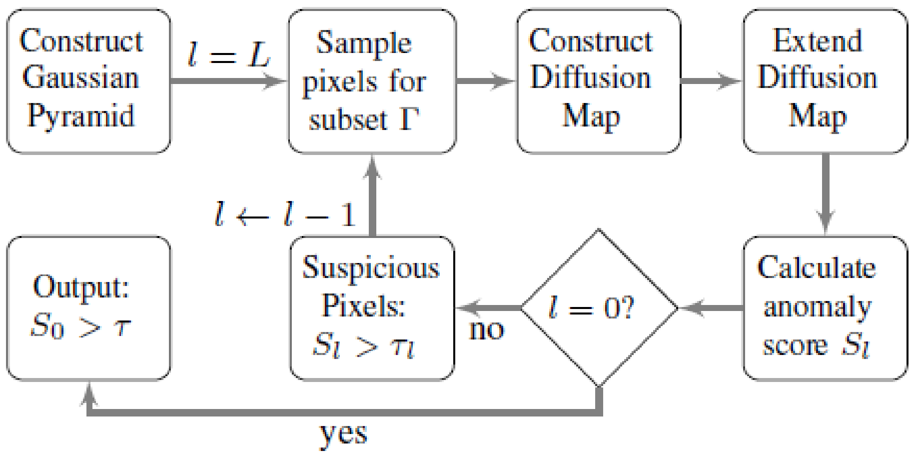


Fig. 1. Flowchart of the multiscale algorithm.

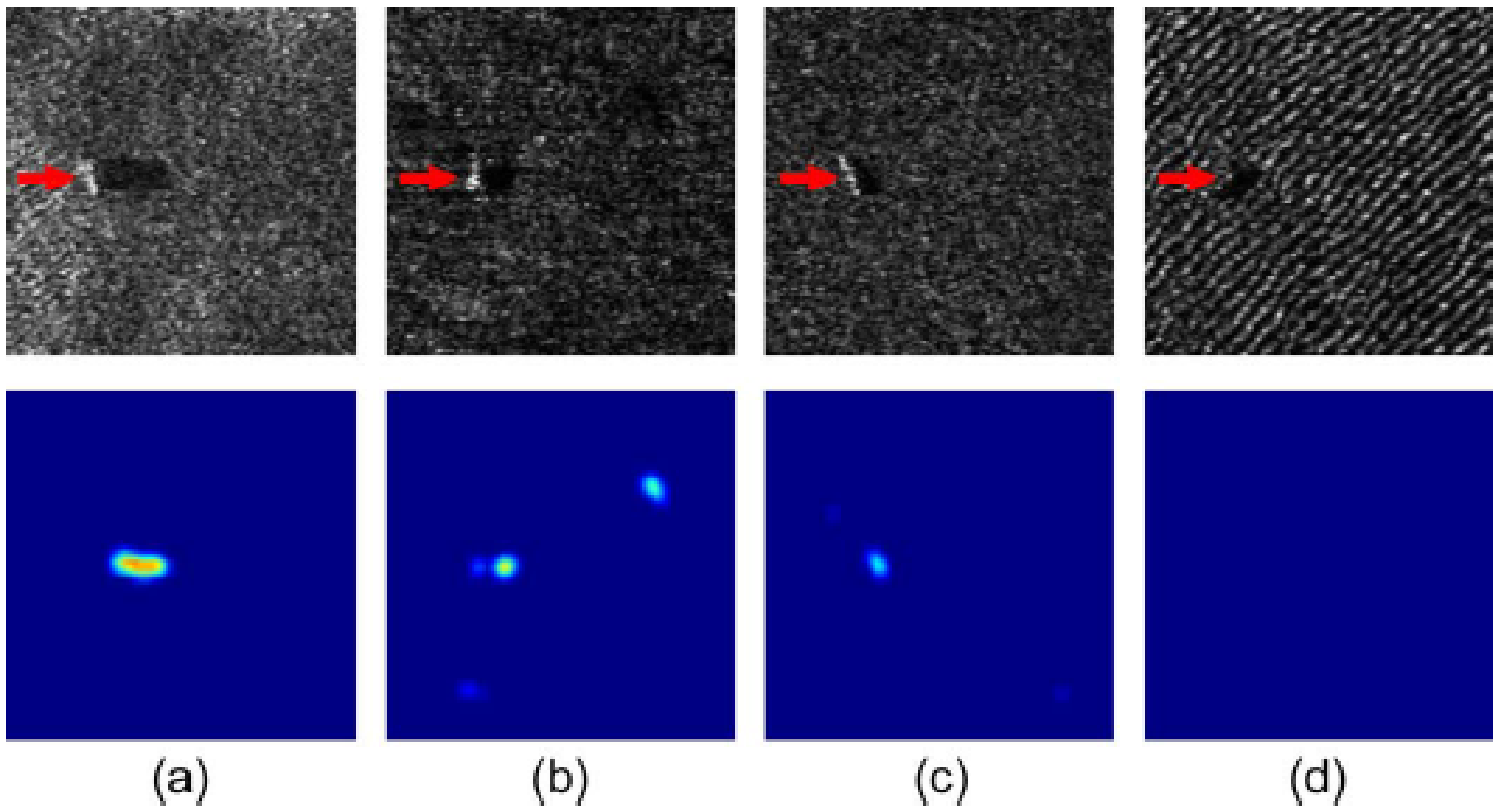


Fig. 2. Top row: original side-scan sonar images, the sea-mines are indicated by red (white in print) arrows. Bottom row: Anomaly score for detection based on coarse resolution of the images. The images were down-sampled by a factor of 2, and a third of the pixels were sampled in the construction of the diffusion map. In (a) the detection is successful. However, this method may result in false alarms (b), low anomaly score (c) or a miss-detection (d).

Novelty Detection in Images by Sparse Representations

Giacomo Boracchi, Diego Carrera, Brendt Wohlberg

Principle 3 lifting to patches followed by sparse analysis of the normal data

We consider two different formulations of *sparse coding*, namely the estimation of a sparse representation for a specific patch $\mathbf{s}_c$ with respect to a given dictionary $\widehat{\mathrm{D}}$:

The unconstrained problem

$$\widehat{\mathbf{x}}_{c,1} = \underset{\mathbf{x} \in \mathbb{R}^n}{\arg\min}\, J_\lambda(\mathbf{x}, \widehat{\mathrm{D}}, \mathbf{s}_c), \tag{4}$$

where the $J_\lambda(\cdot)$ is a convex loss function defined as

$$J_\lambda(\mathbf{x}, \widehat{\mathrm{D}}, \mathbf{s}_c) = \frac{1}{2}\|\widehat{\mathrm{D}}\mathbf{x} - \mathbf{s}_c\|_2^2 + \lambda\|\mathbf{x}\|_1, \tag{5}$$

and $\lambda > 0$ is a regularization parameter that balances the reconstruction error $\|\widehat{\mathrm{D}}\mathbf{x} - \mathbf{s}_c\|_2^2$, and the sparsity $\|\mathbf{x}\|_1$ of the solution measured by the ℓ^1 norm. There are a number of methods for solving this Basis Pursuit DeNoising (BPDN) [20] problem, including Alternating Direction Method of Multipliers (ADMM) [21].

Here we adopt instead a bivariate anomaly indicator, thus jointly accounting for both the reconstruction error and the sparsity of the approximation given by $\widehat{\mathrm{D}}$. In particular, given a patch $\mathbf{s}_c$, we compute the sparse coding $\mathbf{x}_{c,1}$ solving the BPDN (4) problem, and we define the vector

$$\mathbf{g}(\mathbf{s}_c) = [\|\widehat{\mathrm{D}}\mathbf{x}_{c,1} - \mathbf{s}_c\|_2, \|\mathbf{x}_{c,1}\|_1], \tag{8}$$

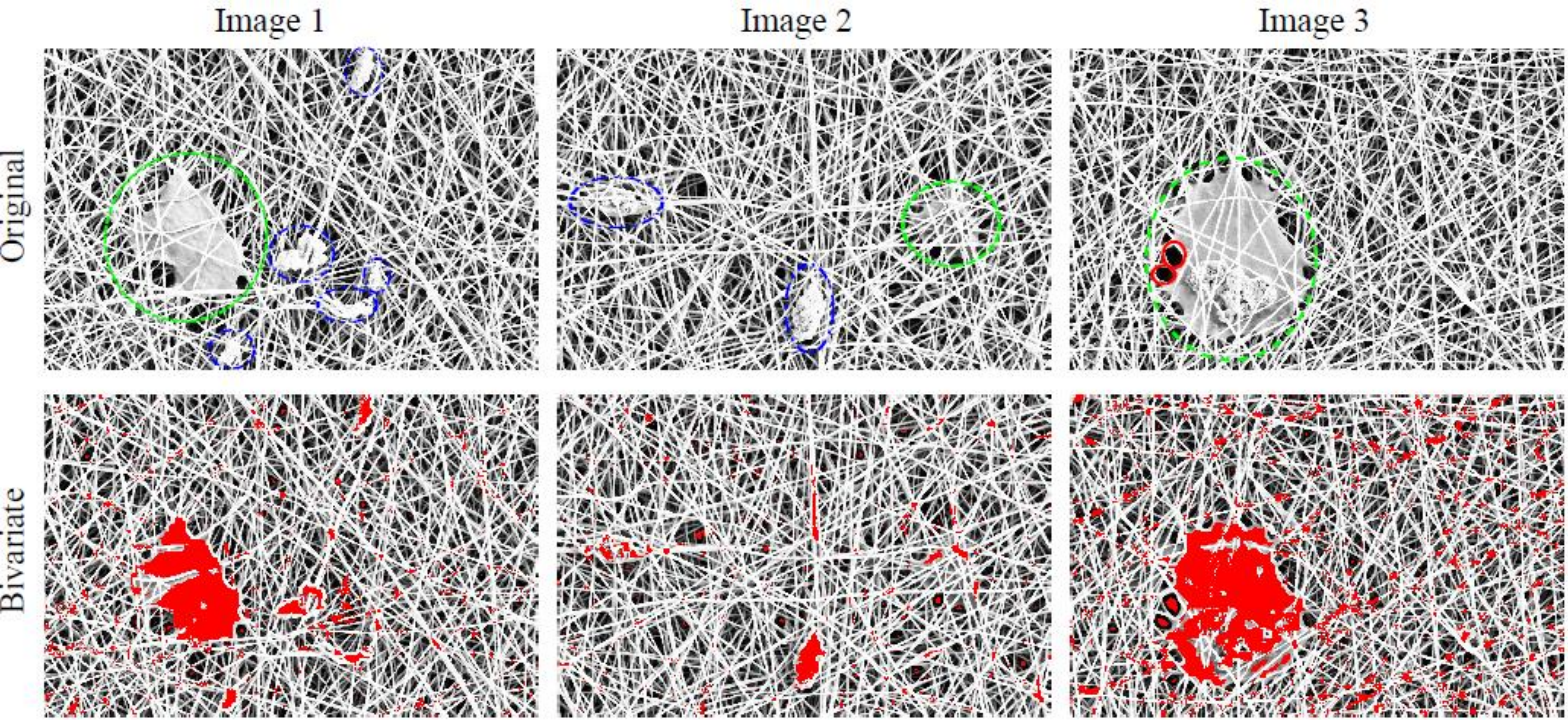
as the bivariate anomaly indicator.

When the bivariate indicator (8) is used, we can build a two-dimensional region [24]

$$\mathcal{R}_\gamma = \left\{\phi \in \mathbb{R}^2 \ : \ \sqrt{(\phi - \mu)^T \Sigma^{-1} (\phi - \mu)} \le \gamma \right\}, \tag{12}$$

where μ and Σ are the expectation and the covariance matrix of $\mathcal{F}$, respectively, and γ is a suitably chosen threshold. Then, a patch $\mathbf{s}_c$ is considered anomalous when it does not belong to $\mathcal{R}_\gamma$, i.e.,

$$\sqrt{(\mathbf{g}(\mathbf{s}_c) - \mu)^T \Sigma^{-1} (\mathbf{g}(\mathbf{s}_c) - \mu)} > \gamma. \tag{13}$$



Clot detection in nanofibers

Novelty Detection in Images by Sparse Representations

Giacomo Boracchi, Diego Carrera, Brendt Wohlberg

What Makes a Patch Distinct?

Ran Margolin Ayellet Tal Lihi Zelnik-Manor

Principle 3 again : sparsity, dimension reduction

Principle 4: locality of detection (image patches)

Mathematically, this boils down to calculating the L_1 norm of p_x in PCA coordinates. Thus, pattern distinctness

$P(p_x)$ is defined as:

$$P(p_x) = ||\tilde{p_x}||_1, \tag{2}$$

where $\tilde{p_x}$ is p_x's coordinates in the PCA coordinate system.

We seek regions that are salient in both color and pattern. Therefore, to integrate color and pattern distinctness we simply take the product of the two:

$$D(p_x) = P(p_x) \cdot C(p_x). \tag{4}$$

This map is normalized to the range $[0, 1]$.

To take these observations under consideration, we do the following. We start by detecting the clusters of distinct pixels by iteratively thresholding the distinctness map $D(p_x)$ using 10 regularly spaced thresholds between 0 and 1. We compute the center-of-mass of each threshold result and place a Gaussian with $\sigma = 10000$ at its location. We associate with each of these Gaussians an importance weight, corresponding to its threshold value. In addition, to accommodate for the center prior, we further add a Gaussian at the center of the image with an associated weight of 5. We then generate a weight map $G(p_x)$ that is the weighted sum of all the Gaussians.

Our final saliency map $S(p_x)$ is a simple product between the distinctness map and the Gaussian weight map:

$$S(p_x) = G(p_x) \cdot D(p_x). \tag{5}$$

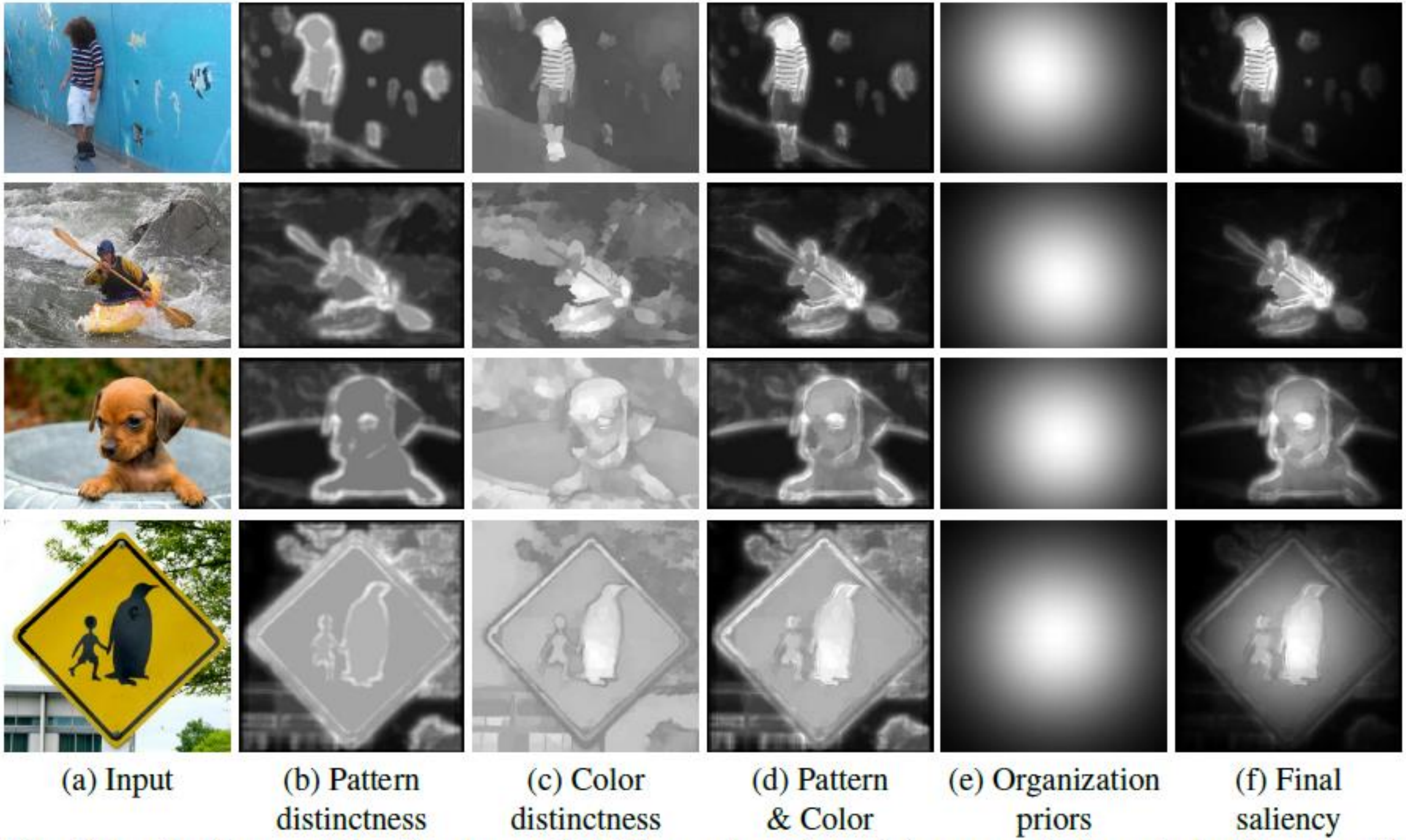


Figure 8. **Combining the three considerations is essential:** Given an input image (a), we compute for each pixel its pattern distinctness (b) and its color distinctness (c). The two distinctness maps are combined (d) and then integrated with priors of image organization (e), to obtain our final saliency results in (f). As can be seen, the final saliency maps are more accurate than each of the components.

Principle 5: Control the number of tests (otherwise you will see « crabs on Mars »)

Principle 6 : evaluate the model by variational method (involving SPARSITY and NOISE)

SCALE-INVARIANT ANOMALY DETECTION WITH MULTISCALE GROUP-SPARSE MODELS
Diego Carrera Giacomo Boracchi Alessandro Foi Brendt Wohlberg ICIP 2016

For simplicity, in the following we illustrate the proposed solution assuming a single training image s is provided, even though multiple training images can be easily handled. **Our solution is based on a dictionary D which is able to approximate any patch taken from an anomaly-free image** as s=Dx; where the coefficients vector x sparse, i.e. has few nonzero or non-negligible components. The dictionary D_i where i corresponds to various scales is learnt by

$$D_i = \underset{D,X}{\arg\min} \frac{1}{2} \|T_i - DX\|_2^2 + \lambda \|X\|_1 \ , \qquad (2)$$

Given the dictionary the best estimate of a patch x is given by

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\arg\min} \frac{1}{2} \|\mathbf{s} - D\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 + \xi \sum_{j=1}^{L} \|\mathbf{x}_j\|_2 \ . \qquad (5)$$

Where the last term is to force the decomposition to choose few scales.

Principle 7 : Define confidence regions or intervals for the « normal » patches

To detect whether a patch s is normal or anomalous, we build a Gaussian confidence region R from the three values of the terms in (5) computed from the normal patches in the training set:

$$\mathcal{R}_\gamma = \left\{ \phi \in \mathbb{R}^3 \ : \ \sqrt{(\phi - \overline{\mathrm{g}})' \Sigma^{-1} (\phi - \overline{\mathrm{g}})} \leq \gamma \right\}, \qquad (7)$$

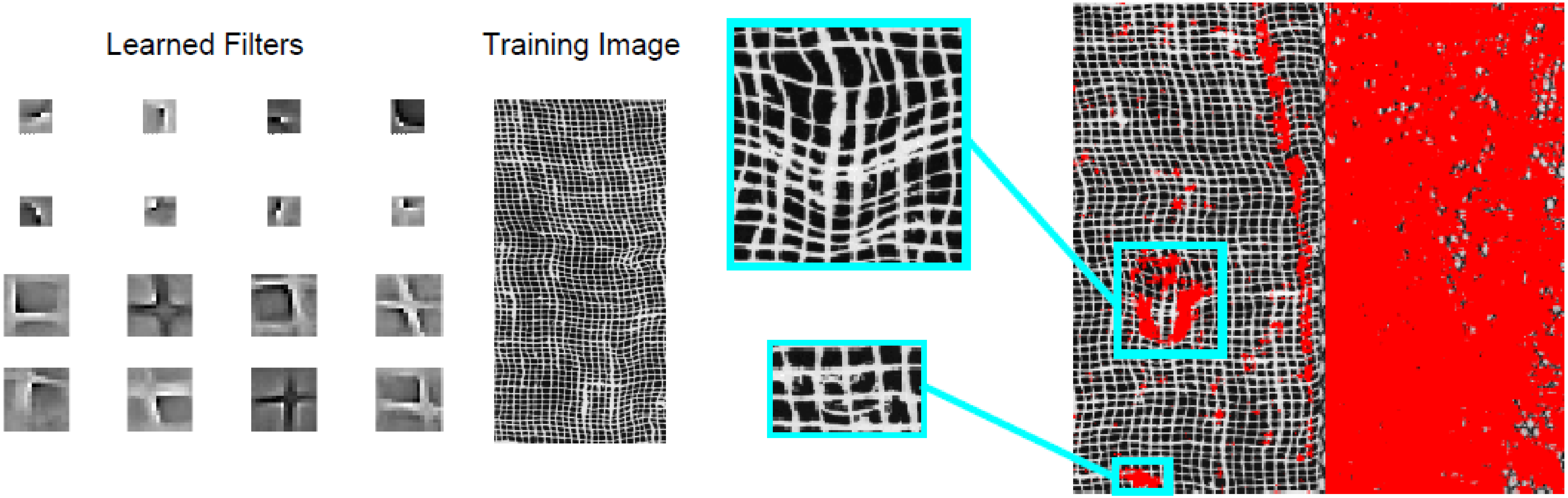


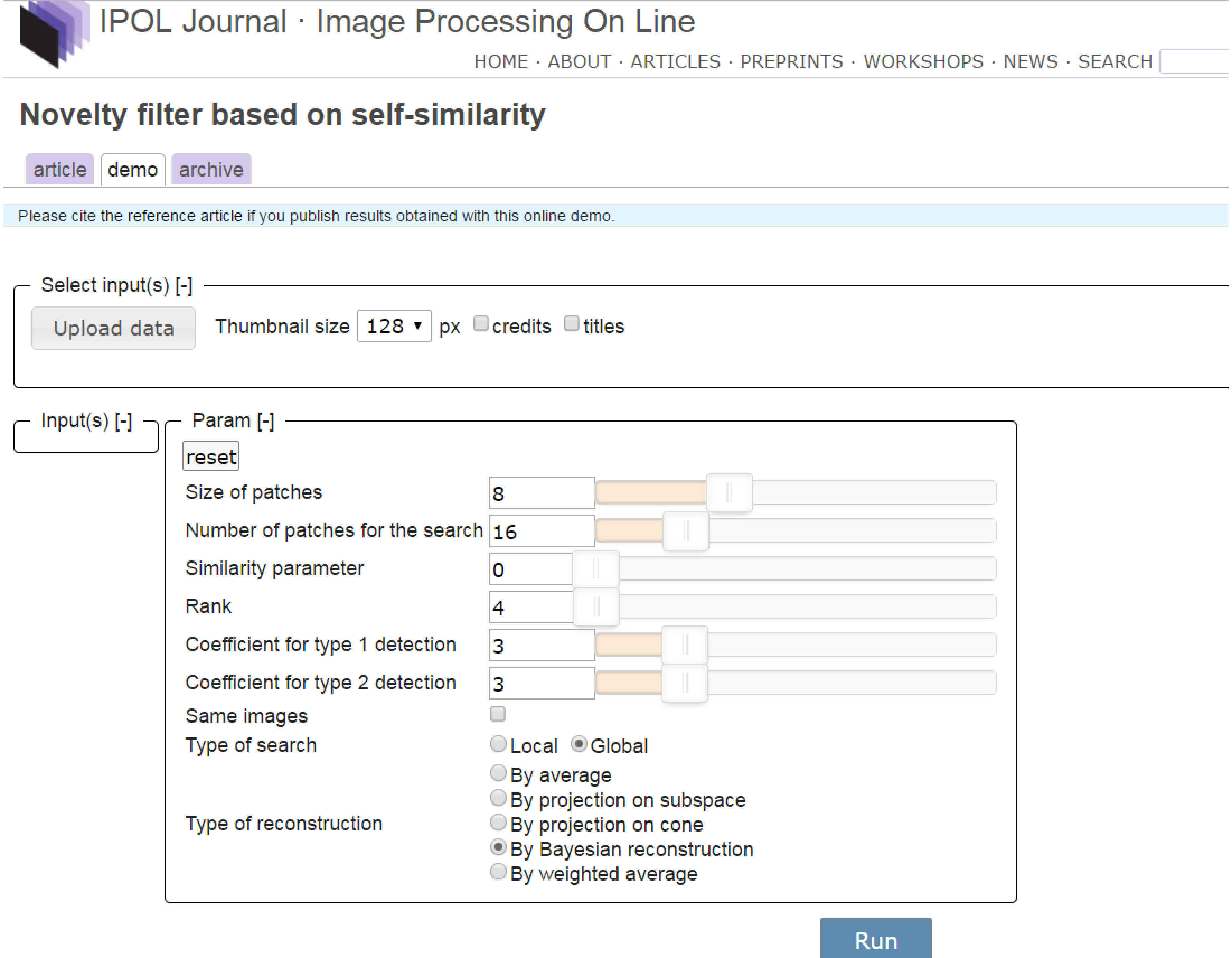
Fig. 4: Example of anomaly-detection performance for the Convolutional Group algorithm. Any detection (red pixels) on the left half represents a false positive, while any detection on the right half a true positive. The ideal anomaly detector would here detect all the points in the left half and none on the right half. Patches across the vertical boundary are not considered in the anomaly detection to avoid artifacts. As shown in the highlighted regions, most of false positives in this example are due to structure that do not conform to the normal image in Figure 2(a).

SCALE-INVARIANT ANOMALY DETECTION WITH MULTISCALE GROUP-SPARSE MODELS

Diego Carrera Giacomo Boracchi
Alessandro Foi Brendt Wohlberg
ICIP 2016

Our general anomaly detection tool

Thibaud Ehret's facility: testing all methods online



Building the image model
(the parameter values are realistic examples)

- Decompose the image u into all of its 8×8 patches P
- Find for each patch P the 16 **most similar** patches P_i, $i = 1, ..., 16$ (located elsewhere in the image)
- Find the **best estimate** of P from the P_i according to one of :
  1. **Mean**
  2. **Mean weighted by kernel-type distance**
  3. **Projection** of P on **Span**$(P_1, \cdots, P_{16})$
  4. **sparse Projection** of P on the **positive cone** generated by $P_1, \cdots, P_{16}$
  5. Bayesian estimate of P given the P_i and a "noise" standard deviation
- Reconstruct an **normal image model** $\tilde{u}$ by aggregating all patch estimates (a simple mean)
- Compute the **difference containing noise + anomalies** $N := \tilde{u} - u$

Type 1 Anomaly detection

- Compute the standard deviation σ of the **noise difference** $N := \tilde{u} - u$. It will be treated as white noise in the *a contrario* model
- Detect all **exceptional pixels** x, such that $\mathbb{P}(N(x)) > s\sigma$, (s=4)
- For $k \in \mathbb{N}$ find all 4-connected components with k exceptional pixels
- Compute the **Number of false alarms** of the **exceptional connected component**,

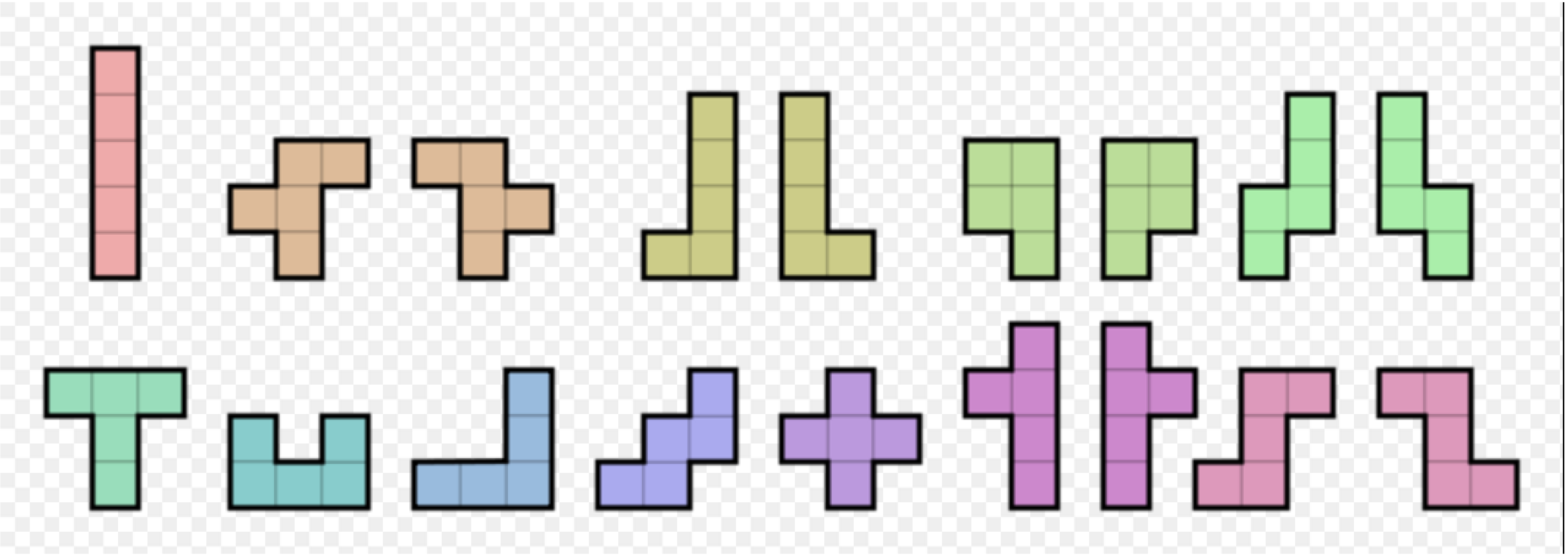
$$NFA(k, s) := N(k).\mathbb{P}(N(x) > s\sigma)^k$$

Desolneux, Agnes, Lionel Moisan, and Jean-Michel Morel. *From gestalt theory to image analysis: a probabilistic approach*. Vol. 34. Springer Science & Business Media, 2007.

$$NFA(k,s) := N(k).\mathbb{P}(N(x) > s\sigma)^k$$

N(k)= number of k polyominos

Polyominos are 4-connected unions of square on regular grid. No closed form seems available for the number of polyominos of size n. In 2004, Iwan Jensen computed the number of fixed size polynomials up to *n* = 56 : For n=56, this number of approximately $6{,}915\times10^{31}$.

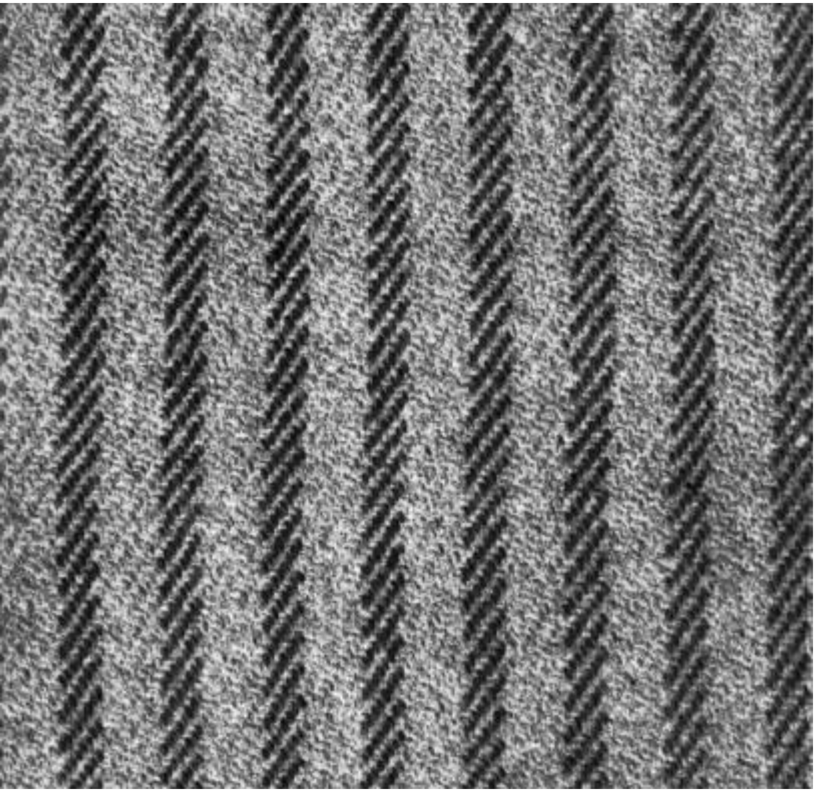


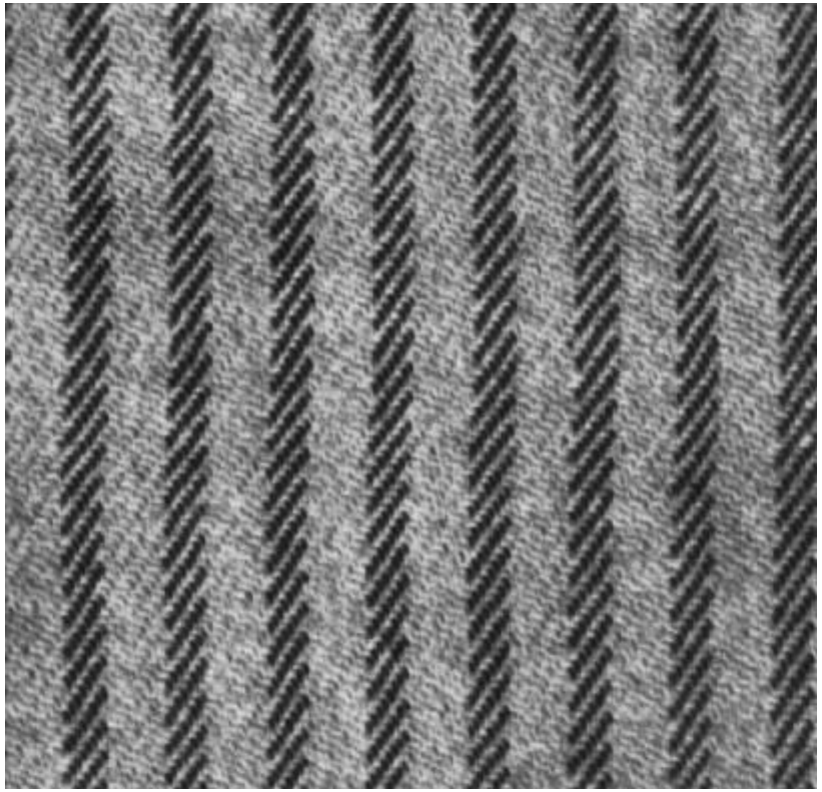
Type 2 Anomaly detection

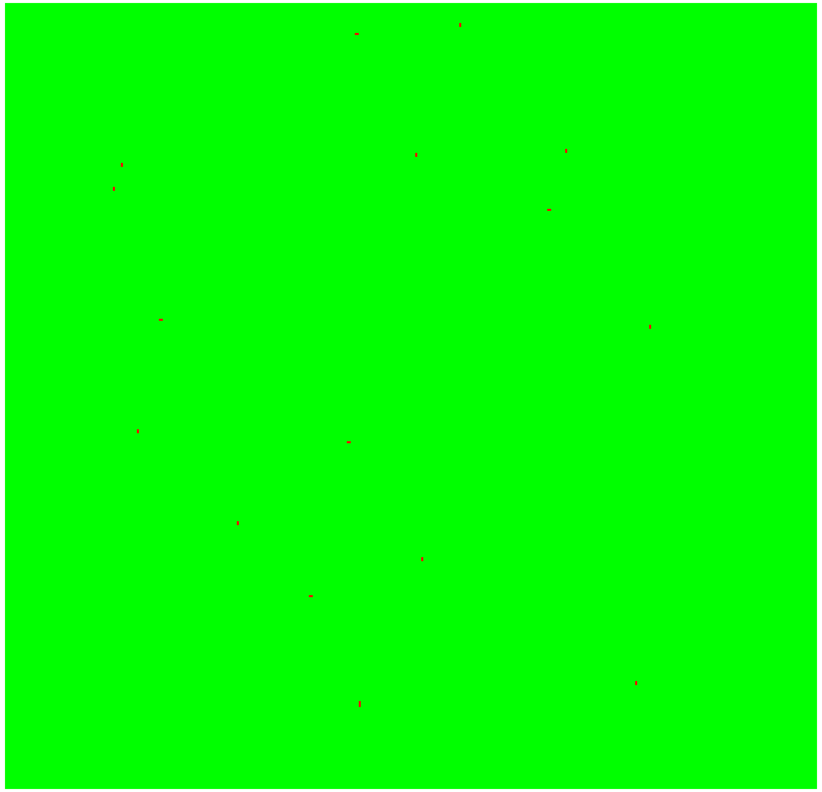
- Take the difference image $N = \tilde{u} - u$ where $\tilde{u}$ is the estimated image model. N should be white noise.
- Compute the standard deviation σ of N
- Detect all exceptional pixels x, such that $\mathbb{P}(N(x)) > s\sigma$, (s=4)
- For each square window W with size n (e.g. $n = 16^2$); count the number k of exceptional pixels in W
- Compute the **Number of false alarms** of the **exceptional square window**, $NFA(k, s) := n' \binom{k}{n} \mathbb{P}(N(x) > s\sigma)^k$. ($n'$ is the number of tested regions)

Desolneux, Agnes, Lionel Moisan, and Jean-Michel Morel. *From gestalt theory to image analysis: a probabilistic approach*. Vol. 34. Springer Science & Business Media, 2007.

Sanity check: very minor detections here



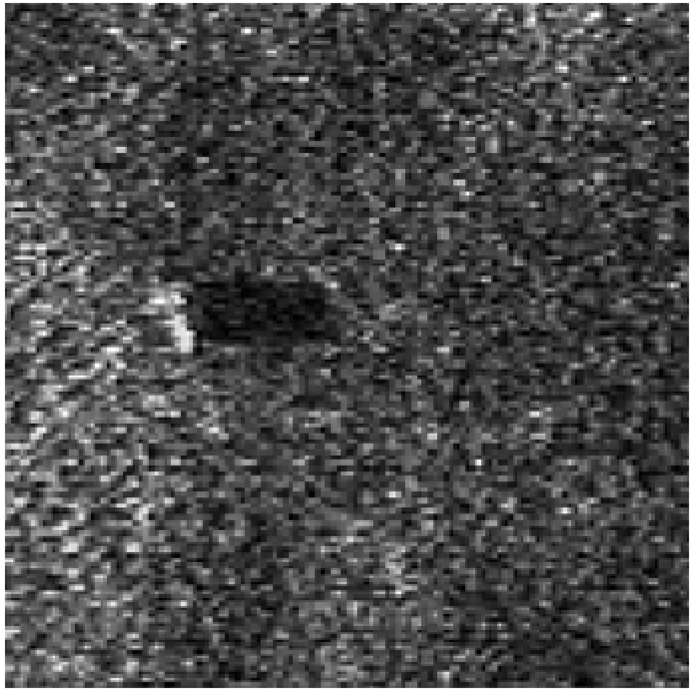


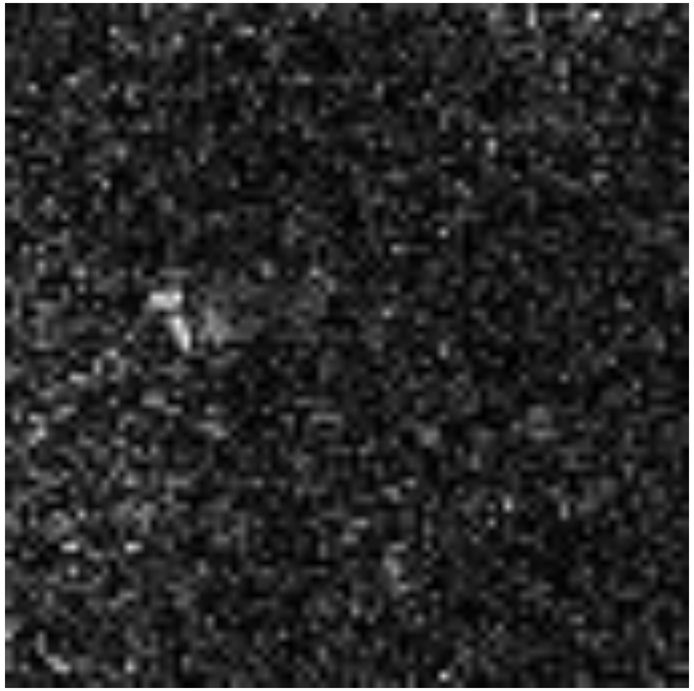


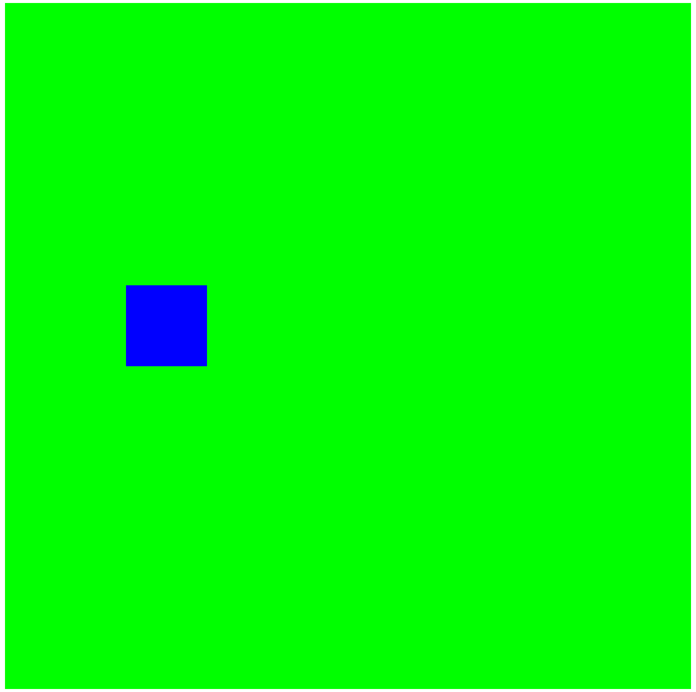
With Bayesian model building:

Raad, L., Desolneux, A., & Morel, J. M. (2015). Conditional Gaussian models for texture synthesis. I*CSS-VM-CV*

Necessity of multiscal detection: side scan sonar detection, only at scale 2







Single image, patch based detection. Example from:
MULTISCALE ANOMALY DETECTION USING DIFFUSION MAPS AND SALIENCY SCORE
Gal Mishne and Israel Cohen

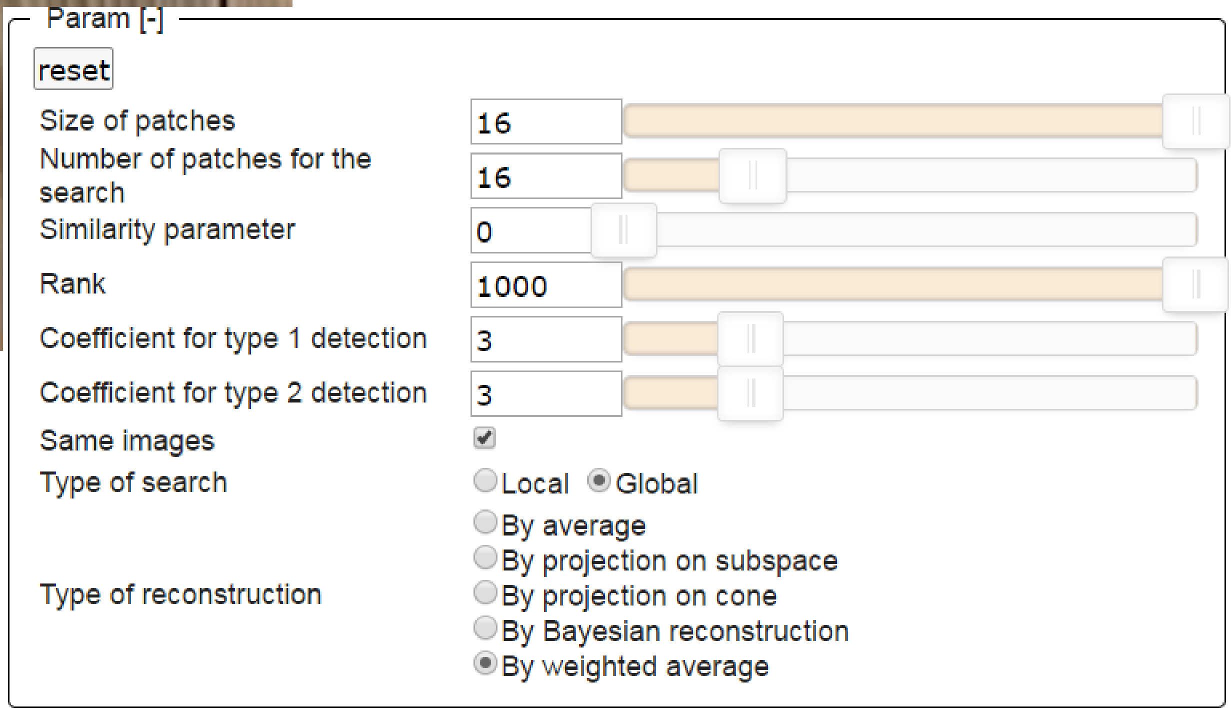
A challenging example: weighted average

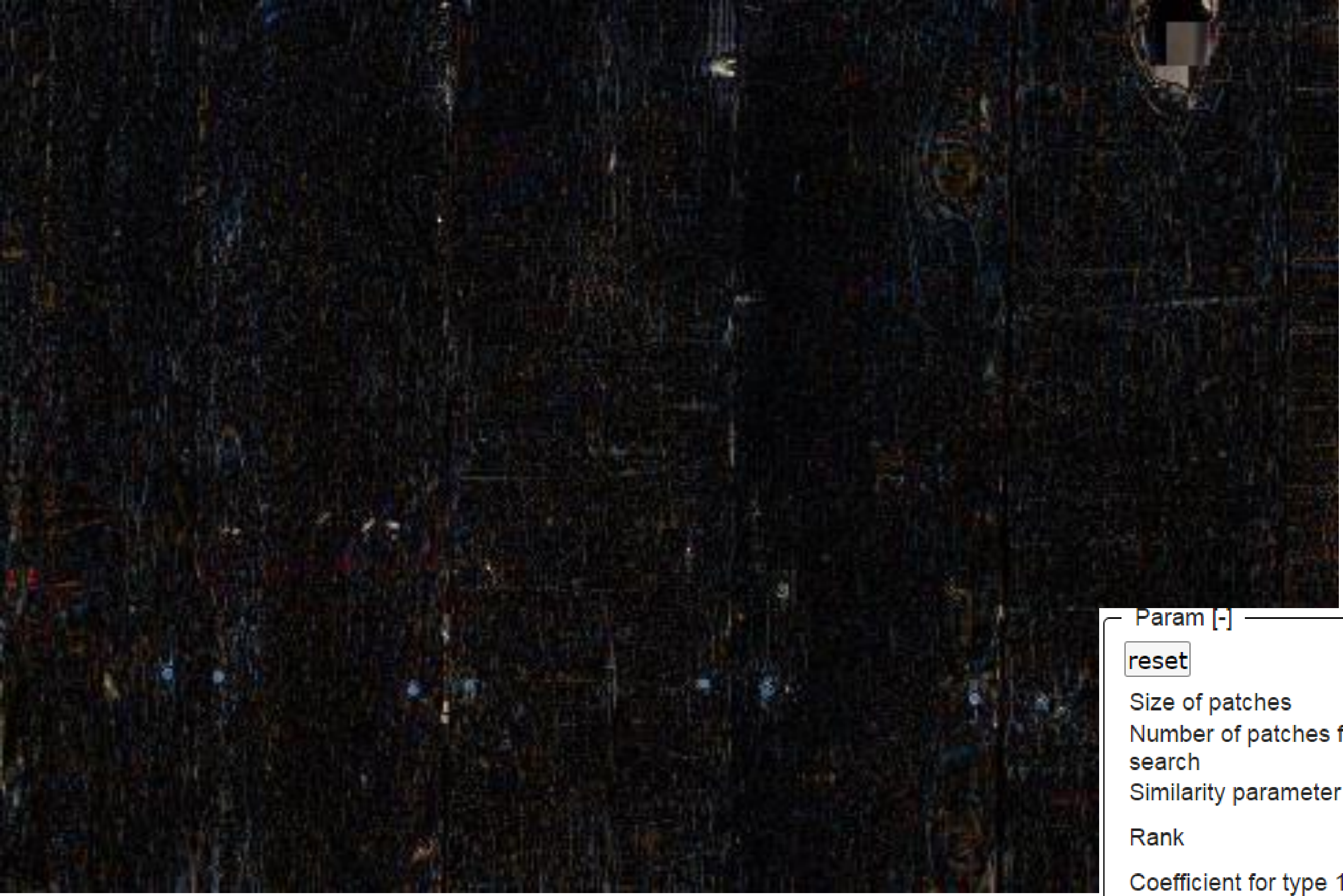
Param [-]

reset

Size of patches	16
Number of patches for the search	16
Similarity parameter	0
Rank	1000
Coefficient for type 1 detection	3
Coefficient for type 2 detection	3
Same images	☑
Type of search	○ Local ◉ Global
Type of reconstruction	○ By average ○ By projection on subspace ○ By projection on cone ○ By Bayesian reconstruction ◉ By weighted average

A challenging example: weighted average



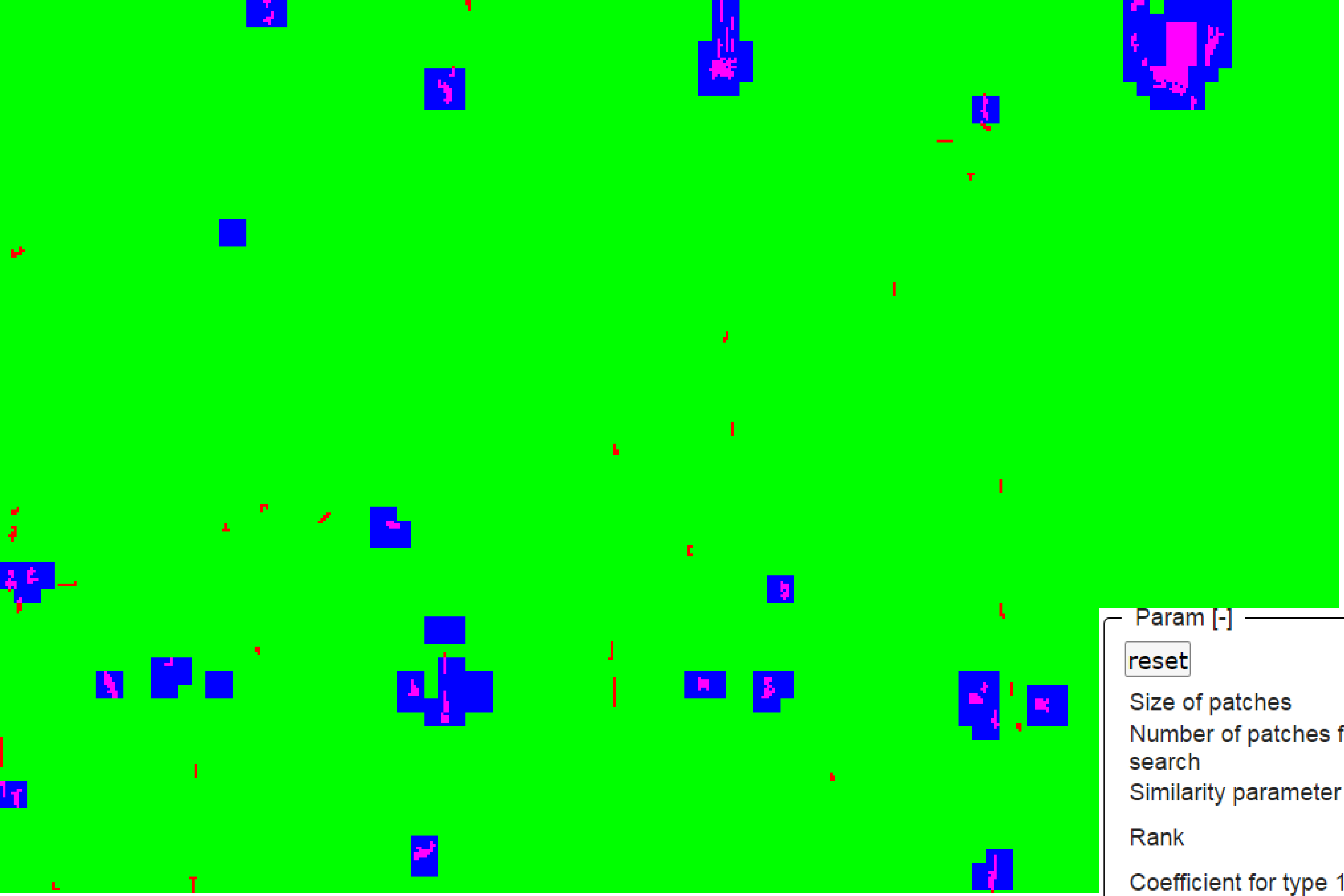


A challenging example: weighted average

Param [-]

reset

Size of patches	16
Number of patches for the search	16
Similarity parameter	0
Rank	1000
Coefficient for type 1 detection	3
Coefficient for type 2 detection	3
Same images	☑
Type of search	○ Local ● Global
Type of reconstruction	○ By average ○ By projection on subspace ○ By projection on cone ○ By Bayesian reconstruction ● By weighted average



A challenging example: weighted average

Param [-]

reset

Size of patches	16
Number of patches for the search	16
Similarity parameter	0
Rank	1000
Coefficient for type 1 detection	3
Coefficient for type 2 detection	3
Same images	☑
Type of search	○ Local ◉ Global
Type of reconstruction	○ By average ○ By projection on subspace ○ By projection on cone ○ By Bayesian reconstruction ◉ By weighted average

A challenging example: weighted average

Param [-]

reset

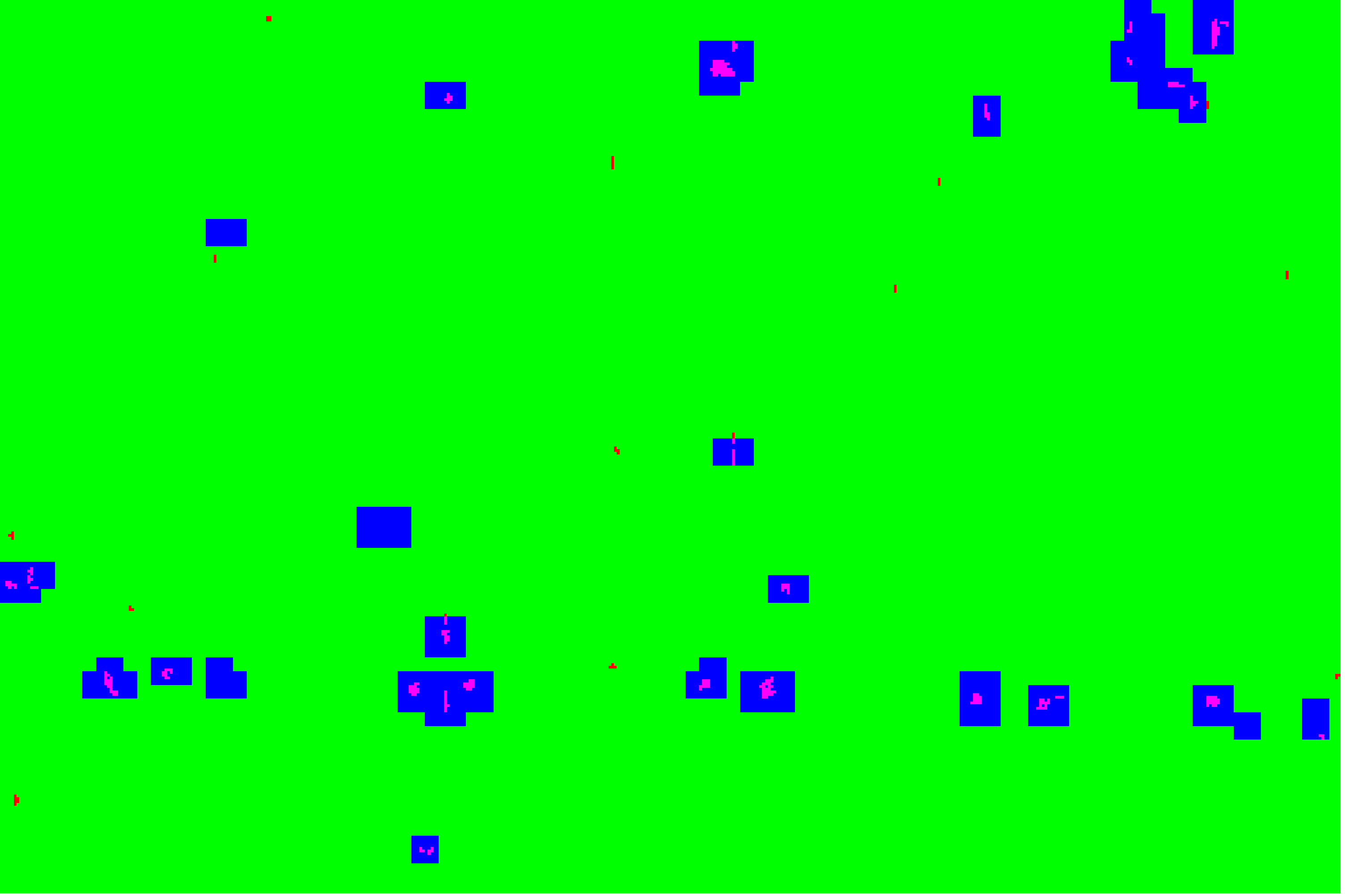
Size of patches	16
Number of patches for the search	16
Similarity parameter	0
Rank	1000
Coefficient for type 1 detection	3
Coefficient for type 2 detection	3
Same images	☑
Type of search	○ Local ◉ Global
Type of reconstruction	○ By average ○ By projection on subspace ○ By projection on cone ○ By Bayesian reconstruction ◉ By weighted average

A challenging example: projection on cone (sparse model decomposition)

Size of patches	16
Number of patches for the search	15
Similarity parameter	0
Rank	8
Coefficient for type 1 detection	4
Coefficient for type 2 detection	4
Same images	☑
Type of search	○ Local ◉ Global
Type of reconstruction	○ By average ○ By projection on subspace ◉ By projection on cone ○ By Bayesian reconstruction ○ By weighted average

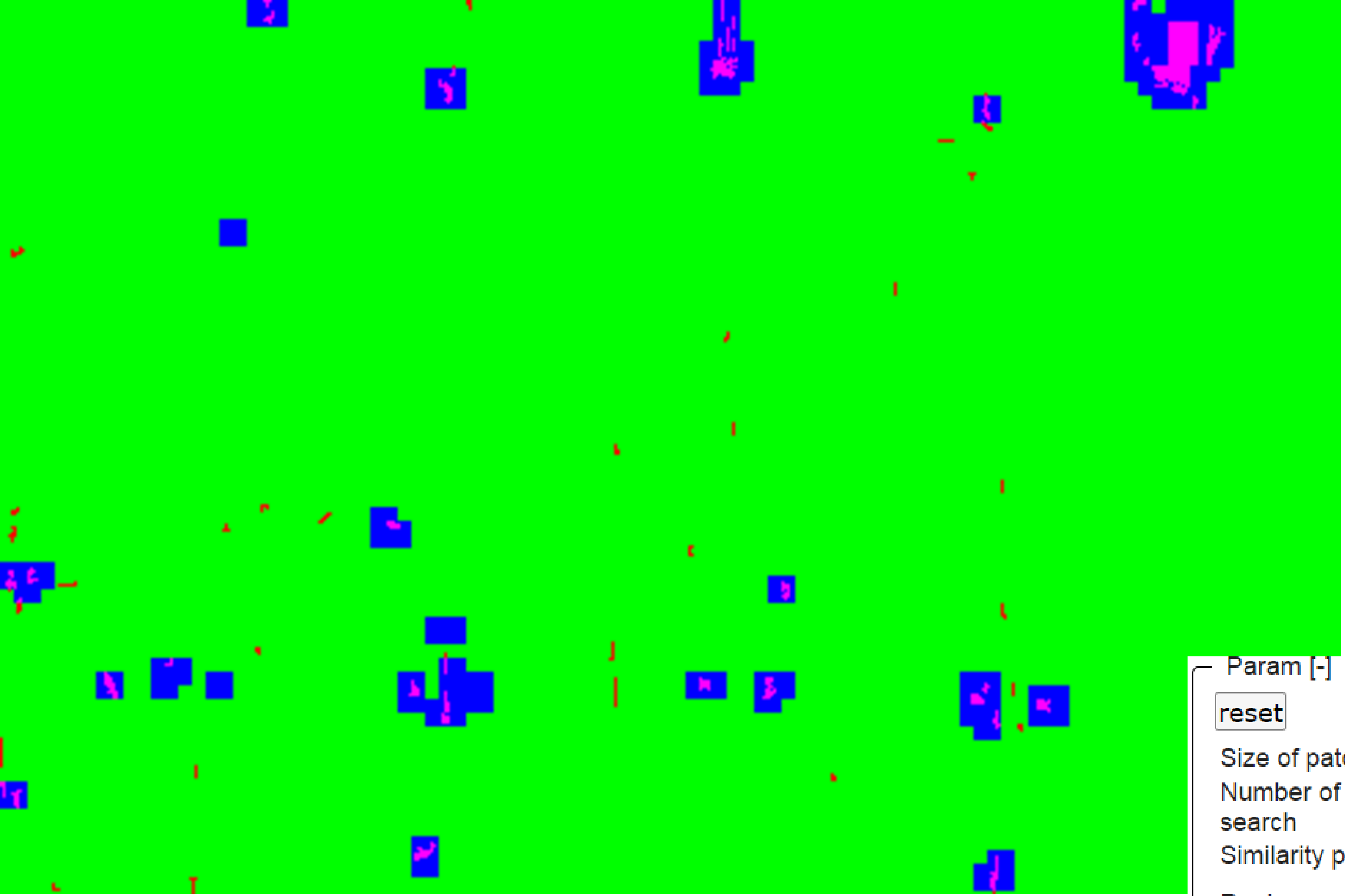
A challenging example: projection on cone

Size of patches	16
Number of patches for the search	15
Similarity parameter	0
Rank	8
Coefficient for type 1 detection	4
Coefficient for type 2 detection	4
Same images	☑
Type of search	○ Local ● Global
Type of reconstruction	○ By average ○ By projection on subspace ● By projection on cone ○ By Bayesian reconstruction ○ By weighted average



A challenging example: projection on cone

Size of patches	16
Number of patches for the search	15
Similarity parameter	0
Rank	8
Coefficient for type 1 detection	4
Coefficient for type 2 detection	4
Same images	☑
Type of search	○ Local ◉ Global
Type of reconstruction	○ By average ○ By projection on subspace ◉ By projection on cone ○ By Bayesian reconstruction ○ By weighted average



A challenging example: weigthed average

Param [-]

reset

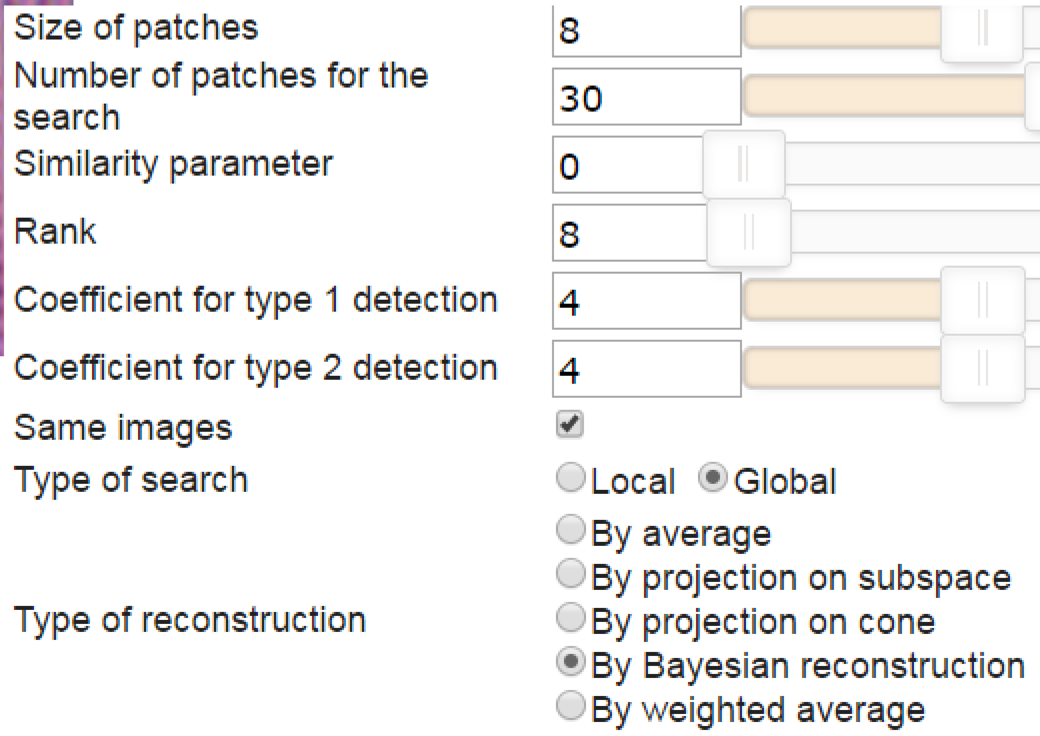
Size of patches	16
Number of patches for the search	16
Similarity parameter	0
Rank	1000
Coefficient for type 1 detection	3
Coefficient for type 2 detection	3
Same images	☑
Type of search	○ Local ◉ Global
Type of reconstruction	○ By average ○ By projection on subspace ○ By projection on cone ○ By Bayesian reconstruction ◉ By weighted average

Bayesian Gaussian detection

Size of patches	8
Number of patches for the search	30
Similarity parameter	0
Rank	8
Coefficient for type 1 detection	4
Coefficient for type 2 detection	4
Same images	☑
Type of search	○ Local ◉ Global
Type of reconstruction	○ By average ○ By projection on subspace ○ By projection on cone ◉ By Bayesian reconstruction ○ By weighted average

Bayesian Gaussian detection

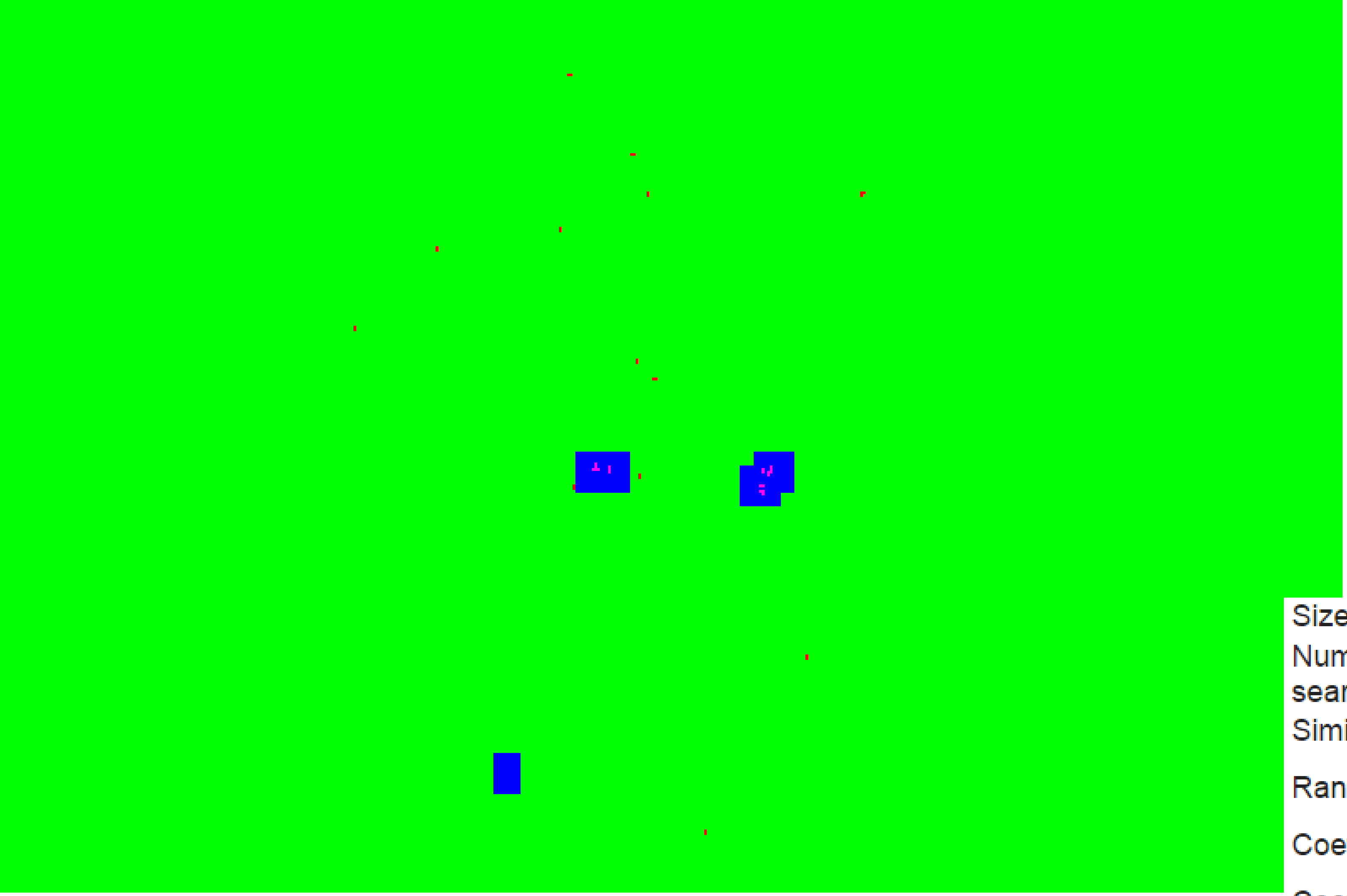
Raad, L., Desolneux, A., & Morel, J. M. (2015). Conditional Gaussian models for texture synthesis. *ICSS-VM-CV*



Bayesian Gaussian detection

Size of patches	8
Number of patches for the search	30
Similarity parameter	0
Rank	8
Coefficient for type 1 detection	4
Coefficient for type 2 detection	4
Same images	☑
Type of search	○ Local ◉ Global
Type of reconstruction	○ By average ○ By projection on subspace ○ By projection on cone ◉ By Bayesian reconstruction ○ By weighted average

Bayesian Gaussian detection



Size of patches	8
Number of patches for the search	30
Similarity parameter	0
Rank	8
Coefficient for type 1 detection	4
Coefficient for type 2 detection	4
Same images	☑
Type of search	○ Local ◉ Global
Type of reconstruction	○ By average ○ By projection on subspace ○ By projection on cone ◉ By Bayesian reconstruction ○ By weighted average

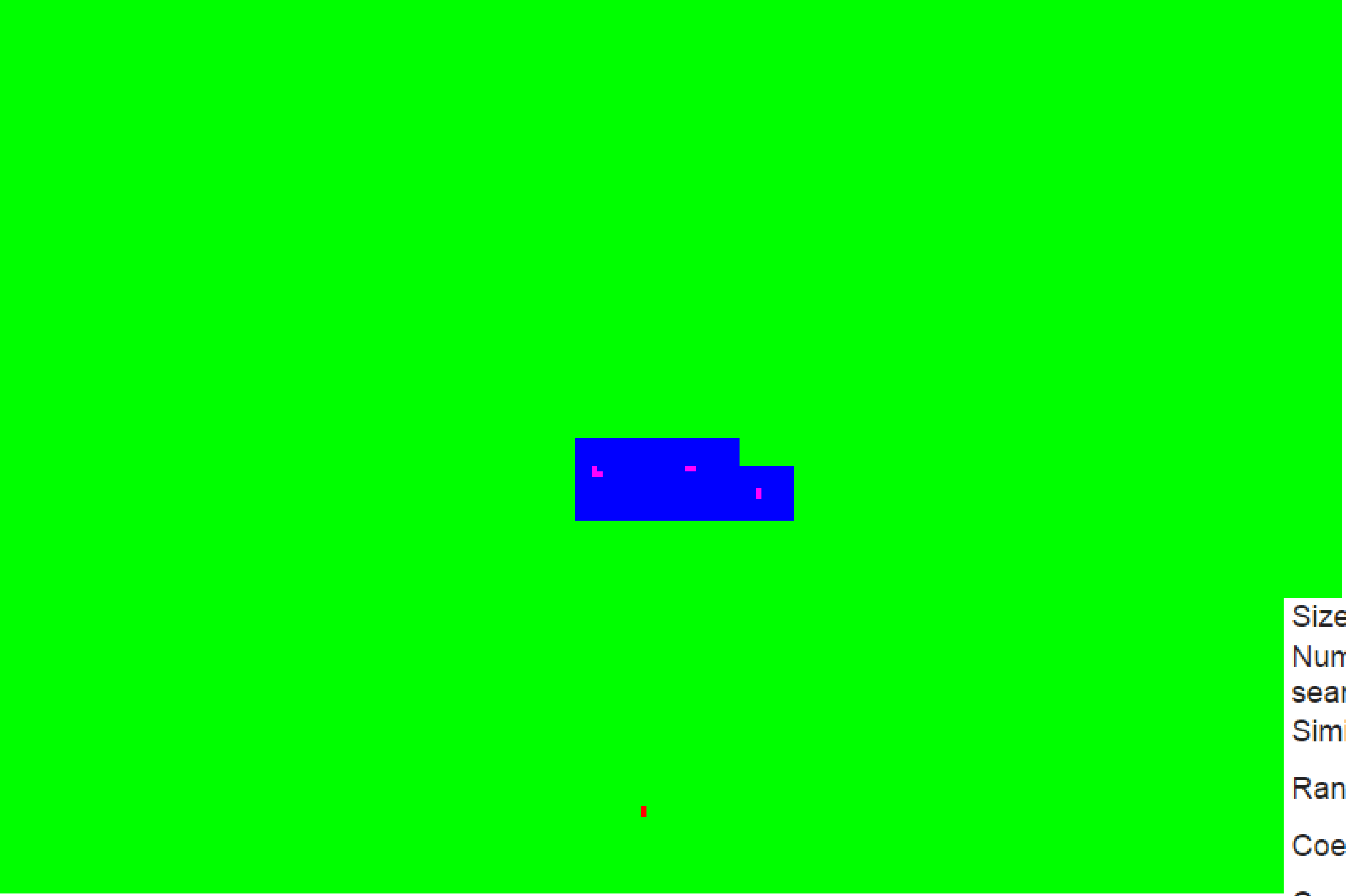
Bayesian Gaussian detection

Size of patches	8
Number of patches for the search	30
Similarity parameter	0
Rank	8
Coefficient for type 1 detection	4
Coefficient for type 2 detection	4
Same images	☑
Type of search	○ Local ◉ Global
Type of reconstruction	○ By average ○ By projection on subspace ○ By projection on cone ◉ By Bayesian reconstruction ○ By weighted average

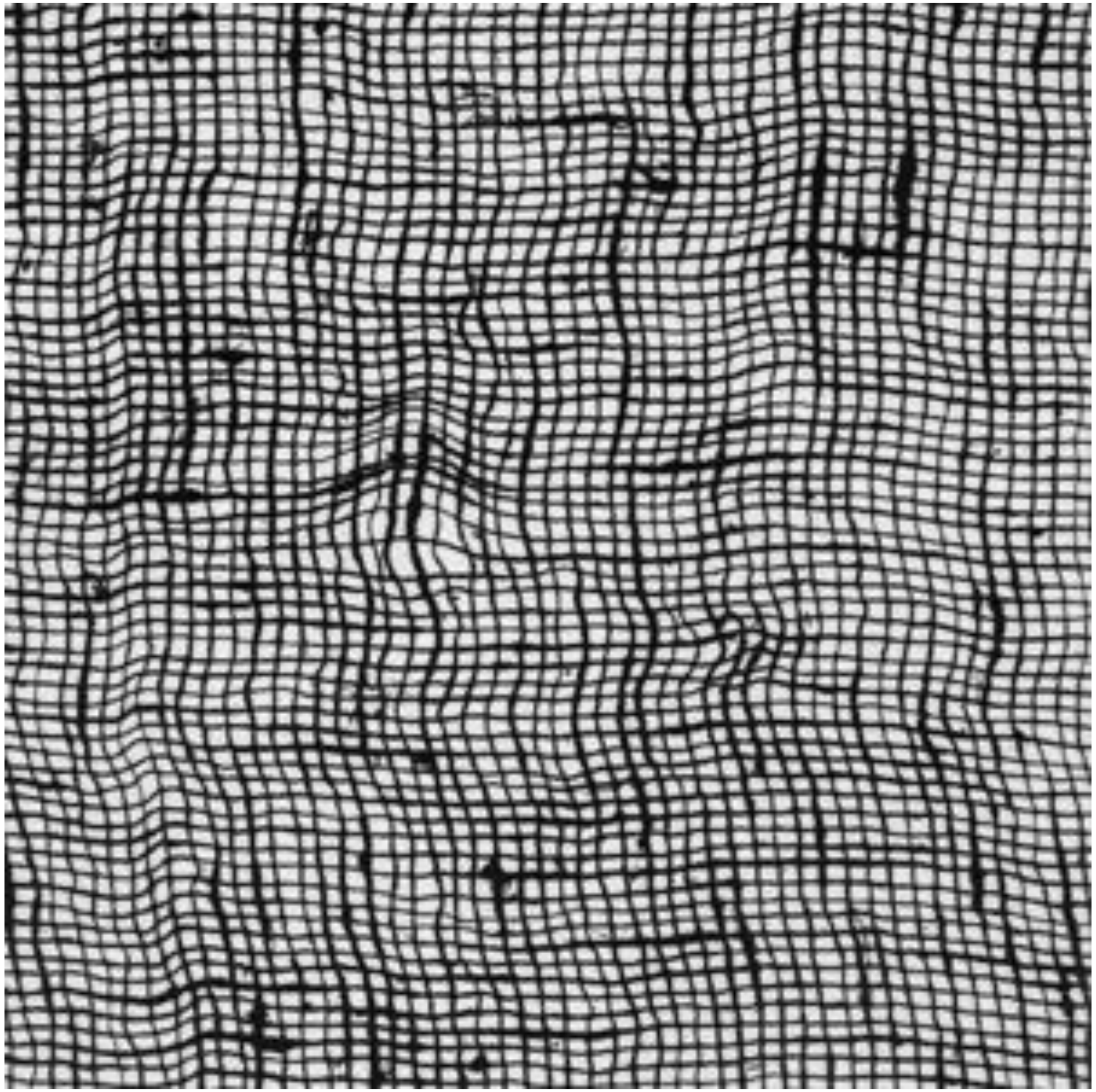
Bayesian Gaussian detection

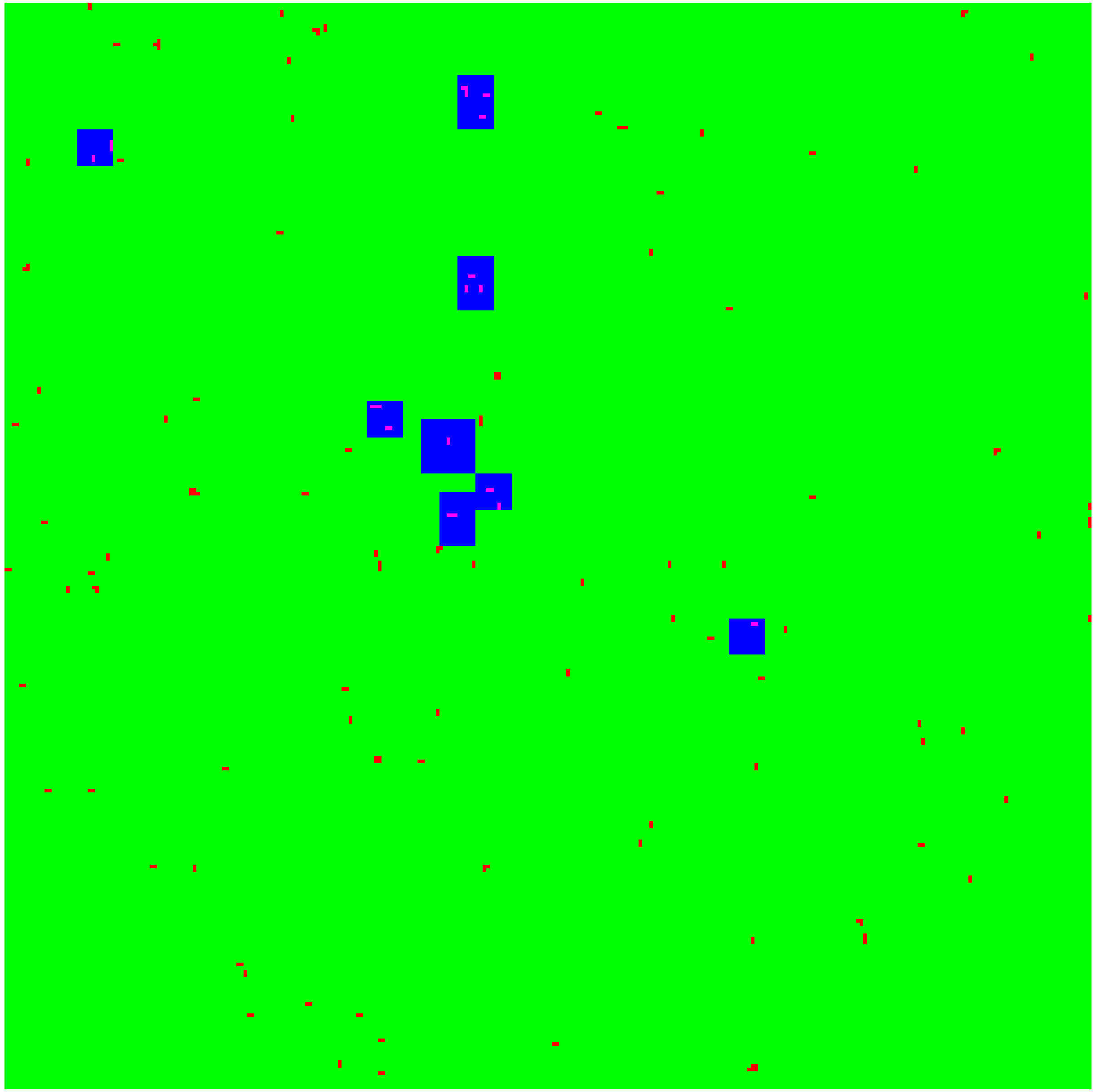
Size of patches	8
Number of patches for the search	30
Similarity parameter	0
Rank	8
Coefficient for type 1 detection	4
Coefficient for type 2 detection	4
Same images	☑
Type of search	○ Local ● Global
Type of reconstruction	○ By average ○ By projection on subspace ○ By projection on cone ● By Bayesian reconstruction ○ By weighted average

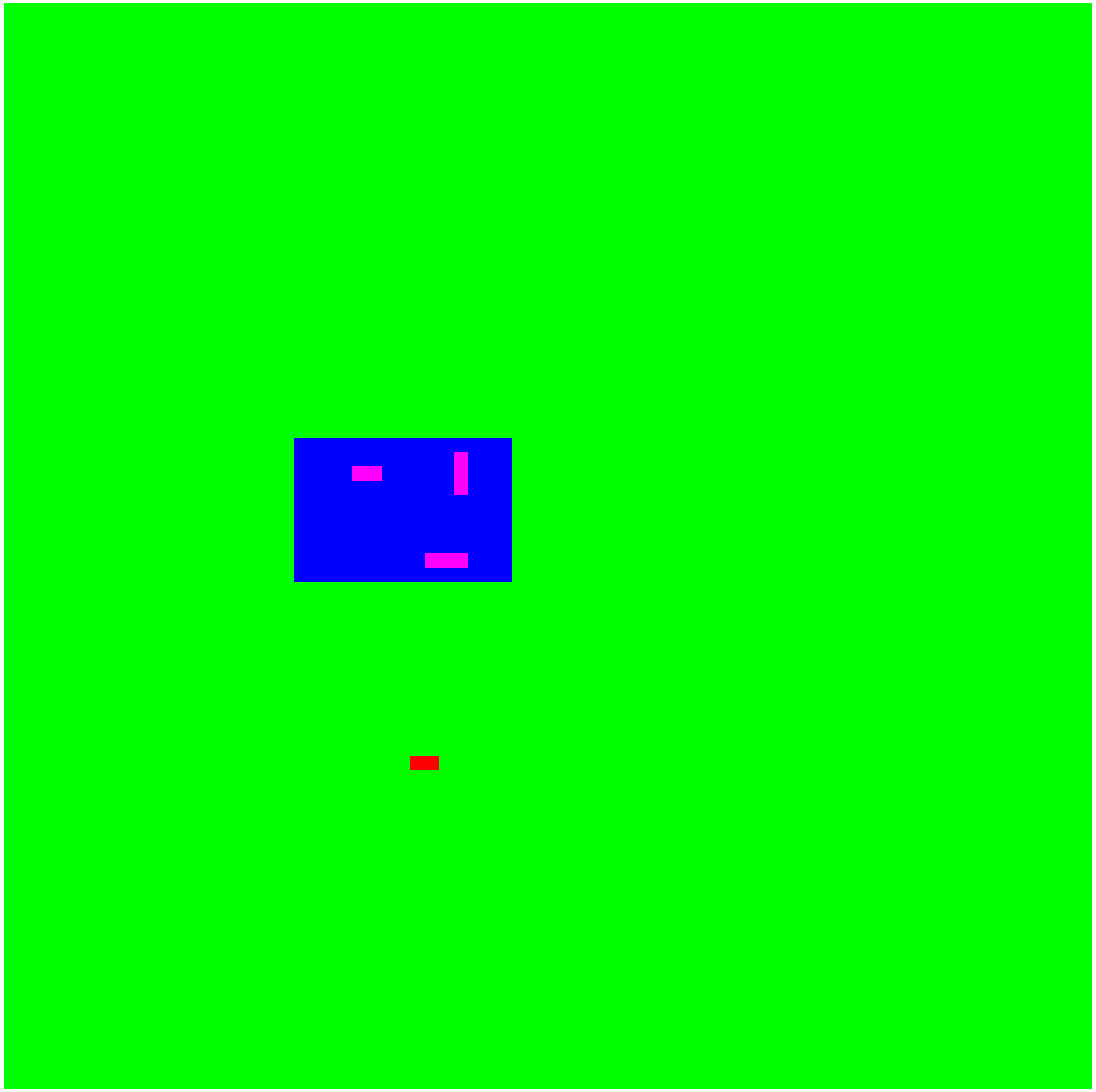
Bayesian Gaussian detection

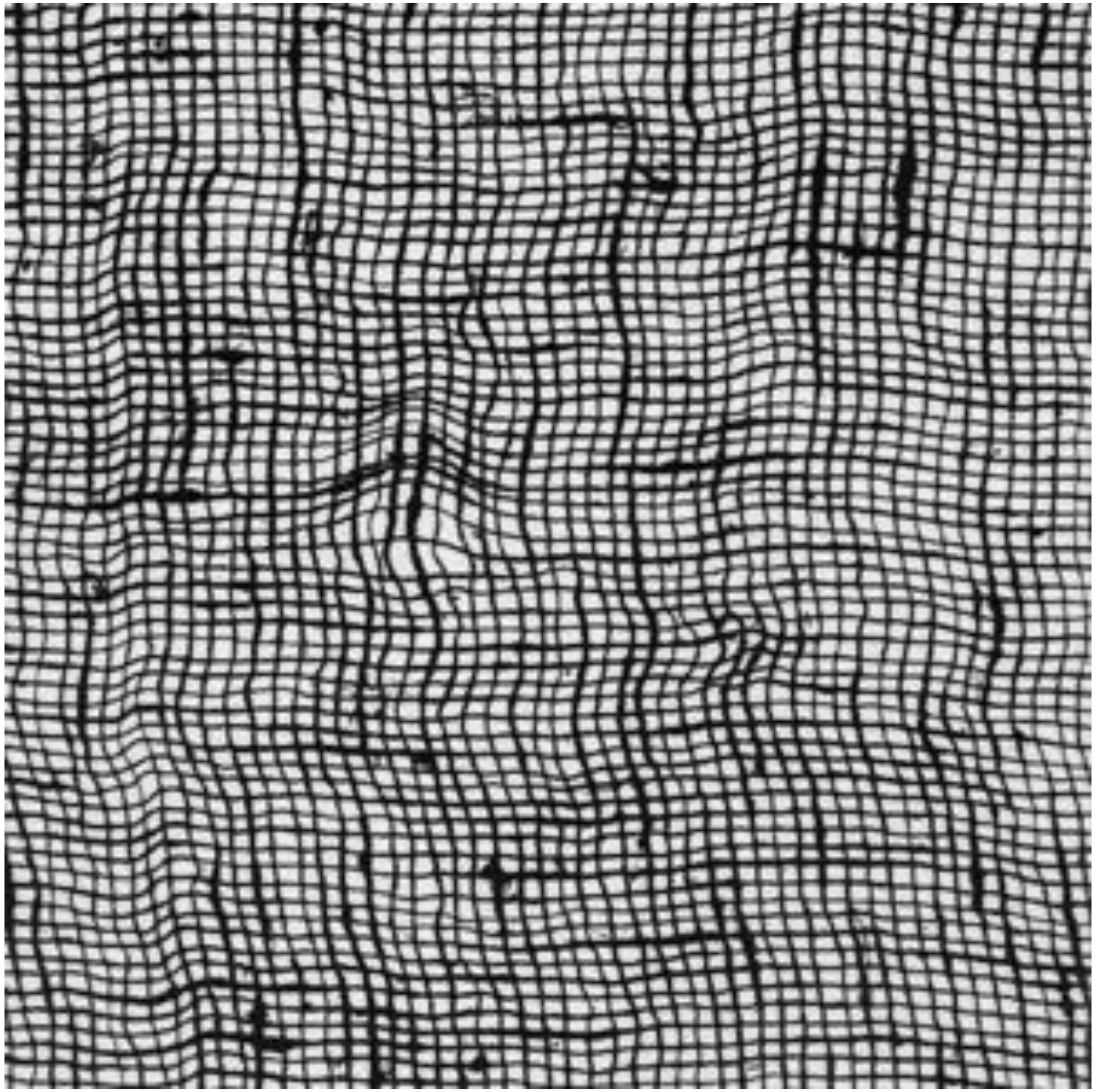


Size of patches	8
Number of patches for the search	30
Similarity parameter	0
Rank	8
Coefficient for type 1 detection	4
Coefficient for type 2 detection	4
Same images	☑
Type of search	○ Local ● Global
Type of reconstruction	○ By average ○ By projection on subspace ○ By projection on cone ● By Bayesian reconstruction ○ By weighted average









Thank you!
Questions?

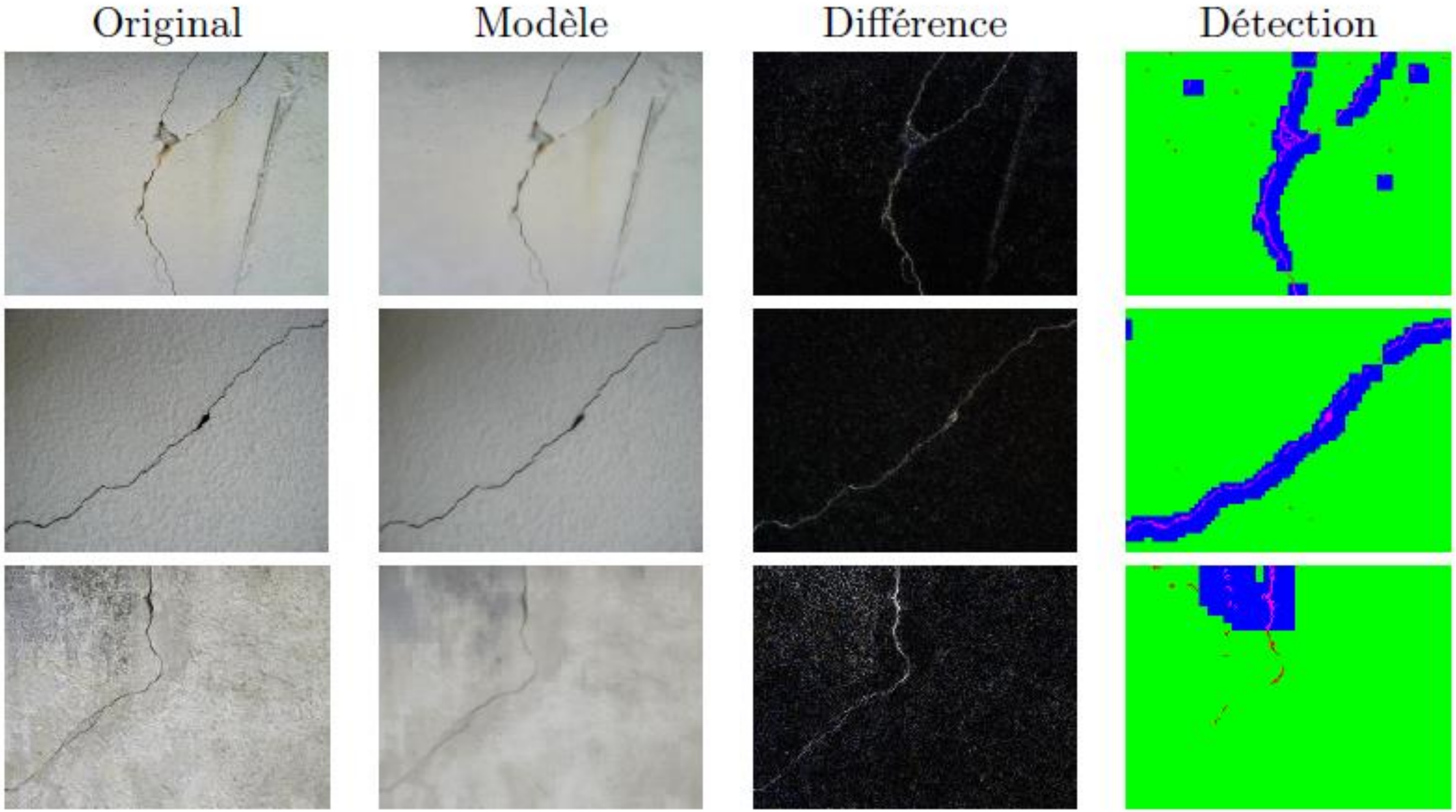


FIGURE 10 – Détections grâce au détecteur d'anomalies sur des images de fissures trouvées sur internet

Four features retained :

- **inter-channel image movement** : application of an algorithm (optical flow) on each pairs of channels of an image, giving a dense displacement map ;
- **image texture characterization** : usage of a robust (to noise, blur, contrast changes) descriptor (SIFT — scale invariant feature transform), describing the spatial gradient distribution in the neighborhood of a keypoint ;
- **inter-image emergence** : usage of a novelty filter highlighting differences of a given image regarding the set of other images ;
- **image brightness** : comparision of the luminance levels in the aeras not yet suspected to be clouds by the 3 others criteria.

Detection theory

For each of the four features, we learn the distribution of the features when no cloud is present, and with the NFA (Number of False Alarms) statistical test, we compare the answers of the features to the modelled distributions, and combine them to create the cloud mask.

Detection theory

demo1+D	http://goo.gl/ta7ktL
demo1+S+D+N+L	http://goo.gl/08Tw91
demo2+S+D+N+L	http://goo.gl/5wngv4
demo3+S+D	http://goo.gl/p4ug9Y
demo3+S+D+N	http://goo.gl/mRNRGI
demo4+S+D	http://goo.gl/1n1YNt
demo4+S+D+N+L	http://goo.gl/67noo0

Given $Y_1, \ldots, Y_n$ Past images, and Z current image.

Basic Novelty filters : Look at the residual after projection on the space of previous images (usually using PCA to limit the space dimension when you have more images than dimensions).

$$\underset{x_1,\ldots,x_n \in \mathbb{R}}{\text{minimize}} \quad ||Z - \sum_{i=1}^{n} x_i Y_i||^2$$

The novelty is $R = Z - \sum_{i=1}^{n} x_i^* Y_i$

Problem : small artifacts can be used to “explain” novelties.

$$\begin{aligned} \underset{x_1,\ldots,x_n \in \mathbb{R}}{\text{minimize}} \quad & ||Z - \sum_{i=1}^{n} x_i Y_i||^2 \\ \text{subject to} \quad & x_i \geq 0 \; i \in [|1, n|], \end{aligned}$$

$\rightarrow$ The quality of the novelty images is improved significantly !

Detecting Anomalous Structures by Convolutional Sparse Models

Diego Carrera, Giacomo Boracchi, Alessandro Foi, Brendt Wohlberg

Dictionary learning is formulated as the following optimization problem

$$\underset{\{d_m\},\{x_m\}}{\arg\min}\ \frac{1}{2}\left\|\sum_{m=1}^{M} d_m * x_m - s_h\right\|_2^2 + \lambda\sum_{m=1}^{M}\|x_m\|_1\,, \tag{4}$$
$$\text{subject to } \|d_m\|_2 = 1,\quad m \in \{1,\dots,M\}\,,$$

where $\{d_m\}$ and $\{x_m\}$ denote the collections of M filters and coefficient maps, respectively.

The indicator based on the high frequency components of the image is defined as

$$\mathrm{g}_h(c) = \begin{bmatrix} \|\Pi_c\,(s_h - \sum_m d_m * x_m)\|_2^2 \\ \sum_m \|\Pi_c x_m\|_1 \\ \sum_m \|\Pi_c x_m\|_2 \end{bmatrix}, \tag{9}$$

2) Sparse Coding: The computation of coefficient maps $\{x_m\}$ of an input image s_h with respect to a dictionary $\{d_m\}$ is referred to as *sparse coding*, and consists in solving the following optimization problem [2]:

$$\underset{\{x_m\}}{\arg\min}\ \frac{1}{2}\left\|\sum_{m} d_m * x_m - s_h\right\|_2^2 + \lambda\sum_{m}\|x_m\|_1\,, \tag{7}$$

where filters $\{d_m\}$ were previously learned from (4).

3) Detecting Anomalous Patches: We treat indicators as random vectors and detect as anomalous all the patches yielding indicators that can be considered outliers. Therefore, we build a confidence region $\mathcal{R}_\gamma$ around the mean vector [20] for normal patches, namely:

$$\mathcal{R}_\gamma = \left\{\phi \in \mathbb{R}^2\ :\ \sqrt{(\phi - \overline{\mathrm{g}})^T\Sigma^{-1}(\phi - \overline{\mathrm{g}})} \le \gamma\right\}, \tag{12}$$

where $\overline{\mathrm{g}}$ and Σ denote the sample mean and sample covariance matrix of indicators extracted from normal images in T, and $\gamma > 0$ is a suitably chosen threshold. $\mathcal{R}_\gamma$ represents an high-density regions for indicators extracted from normal patches,

Detection theory

Given the fact that we can never train a machine learning system on all possible object classes whose data the system is likely to encounter, it becomes important that it is able to differentiate between known and unknown object information during testing. It has been realised in practice by several studies that the novelty detection is an extremely challenging task. . (Novelty detection: a review—part 1: statistical approaches Markos Markou, Sameer Singh)

An assumption is made that the abnormalities are uniformly distributed outside the boundaries of normality. The description of normality is made using the unconditional probability density estimation of the training data. If a test vector falls in a region of input space with a density under a pre-determined threshold then the test vector is considered to be novel. (L. Tarassenko, Novelty detection for the identification of masses in mammograms, Proceedings of the 4th IEE International Conference on Artificial Neural Networks, Vol. 4, Cambridge, UK, 1995, pp. 442–447.

A hyper-sphere is drawn to separate known regions from unknown regions. Novel objects should ideally fall outside this hypersphere. An appropriate threshold separates known from new test objects. (L. Parra, G. Deco, S. Miesbach, Statistical independence and novelty detection with information preserving non-linear maps, Neural Comput. 8 (2) (1995) 260–269.

The nearest neighbor method: The distance of the new object and its nearest neighbour in the training set is found and the distance of this nearest neighbour and its nearest neighbour in the training set is also found. The quotient between the first and the second distance is taken as indication of the novelty of the object. (D.M.J. Tax, R.P.W. Duin, Outlier detection using classifier instability, in: Advances in Pattern Recognition, the Joint IAPR International Workshops, Sydney, Australia, 1998, pp. 593–601.)   See also David Lowe, SIFT method.

In this paper we have presented a survey of novelty detection using statistical approaches. Most of such research is driven by modelling data distributions and then estimating the probability of test data to belong to such distributions. In such model-based approaches, one does need to specify or make assumptions on the nature of training data. In addition, the amount and quality of training data becomes very important in the robust determination of training data distribution parameters. (Novelty detection: a review—part 1: statistical approaches Markos Markou, Sameer Singh)

Anomaly detection based on an iterative local statistics approach

Arnon Goldman, Israel Cohen Signal Processing 2004

Let $\mu = E[q_\lambda|H_0]$ denote the expected feature vector and $\Sigma = E[(q_\lambda - \mu)(q_\lambda - \mu)^{\mathrm{T}}|H_0]$ the covariance matrix under H_0 hypothesis. Let the normalized distance of q_λ from its expected vector, μ, be defined by

$$d(q_\lambda) = (q_\lambda - \mu)^{\mathrm{T}} \Sigma^{-1} (q_\lambda - \mu). \tag{3}$$

Then the decision rule is given by

$$d(q_\lambda) \underset{H_1}{\overset{H_0}{\lesseqgtr}} D, \tag{4}$$

where D is the threshold to determine whether a given pixel is anomalous or not. This decision rule is based on the statistics of the background only. No knowledge about the anomalies statistics is taken into consideration. The threshold, D, can be determined according to a specified confidence level, η, which is the probability of correctly deciding on H_0 given H_0 is true. The threshold, D, and the confidence level, η, are related by

$$\eta \equiv \Pr(H_0|H_0) = \Pr(d(q_\lambda) \leqslant D|H_0). \tag{5}$$

In case the feature vector, q_λ, is a Gaussian random vector of dimension n, the pdf of $d^2(q_\lambda)$ under the H_0 hypothesis, denoted by $p_{d^2}(\zeta)$, is the gamma density function with parameters $\beta = n/2 - 1$ and $\alpha = 1/2$ [5]. Accordingly, the relation between η and D can be written as

$$\eta = \int_0^{D^2} p_{d^2}(\zeta)\,\mathrm{d}\zeta = \int_0^{D^2} \frac{1}{2^{n/2}\Gamma(n/2)} \zeta^{(n-2)/2} \mathrm{e}^{-\zeta/2}\,\mathrm{d}\zeta. \tag{6}$$

This is a clear cut hypothesis testing framework, but the authors do not take into account the number of tests they are making, which may well explain their overdetection. Their probably incorrect interpretation is that their decision rule is not sufficient. Thus they iterate several times the division between background and foreground, each time reestimating the covariance matrix for the background.

DETECTION OF ANOMALIES IN TEXTURES BASED ON MULTI-RESOLUTION FEATURES

Lior Shadhan and Israel Cohen 2006

Compared to the preceding reference, the main difference is the way the descriptor is built.

Let $\{y_j(\mathrm{s})\}_{j=1,\ldots,m}$ denote the jth layer wavelet coefficients obtained from the mean normalized image observations $y(\mathrm{s})$ using a RDWT with $(m-1)/3$ levels. Let $\{t_j(\mathrm{s})\}_{j=1,\ldots,m}$ denote the logarithm of the GSM hidden multipliers estimate, given by:

$$t_j(\mathrm{s}) = \log\left(\frac{\sum_{\mathrm{r}\in\mathcal{R}_1} y_j^2(\mathrm{s}+\mathrm{r})}{|\mathcal{R}_1|}\right), \tag{3}$$

where $\mathcal{R}_1$ denotes a given set of indices representing the $N \times N$ local neighborhood of a pixel. Let $\{v_j(\mathrm{s})\}_{j=1,\ldots,m}$ denote the proposed feature space, given by:

$$v_j(\mathrm{s}) = \frac{\sum_{\mathrm{r}\in\mathcal{R}_2} t_j(\mathrm{s}+\mathrm{r})}{|\mathcal{R}_2|}, \tag{4}$$

where $\mathcal{R}_2$ denotes a given set of indices representing the $M \times M$ local neighborhood of a pixel. Let $\mathbf{v}(\mathrm{s}) = [v_1(\mathrm{s}), v_2(\mathrm{s}), \ldots, v_m(\mathrm{s})]^T$ denote the feature vector representing pixel $\mathrm{s} \in \Omega$. Let $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ denote the expectancy of the random feature vector $\mathbf{v}(\mathrm{s})$ under hypotheses H_0 and H_1 respectively. Let $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_1$ denote the covariance matrix of the random feature vector $\mathbf{v}(\mathrm{s})$ under hypotheses H_0 and H_1 respectively. Following the assumption that the anomalous targets are rare and can be regarded as transients:

$$\begin{aligned} \hat{\boldsymbol{\mu}}_0 &\approx E\left[\mathbf{v}(\mathrm{s})\right] \\ \hat{\boldsymbol{\Sigma}}_0 &\approx E\left[(\mathbf{v}(\mathrm{s}) - \hat{\boldsymbol{\mu}}_0)(\mathbf{v}(\mathrm{s}) - \hat{\boldsymbol{\mu}}_0)^T\right]. \end{aligned} \tag{5}$$

The Mahalanobis distance for pixel $\mathrm{s} \in \Omega$ is then given by:

$$d(\mathrm{s}) = (\mathbf{v}(\mathrm{s}) - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1} (\mathbf{v}(\mathrm{s}) - \boldsymbol{\mu}_0). \tag{6}$$

Following the SHT scheme, the decision rule for pixel $\mathrm{s} \in \Omega$ is defined as follows:

$$d(\mathrm{s}) \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{7}$$

where η is the threshold that determines if a given pixel $\mathrm{s} \in \Omega$ is regarded as an anomaly or background clutter. This deci-

The feature vector $\mathbf{v}(\mathrm{s})$ is a linear combination of Gaussian random vectors with dimension m [10, 11] and as such, it is also a Gaussian random vector. Since the covariance matrix Σ_0 is a positive definite matrix, equation (6) can be formulated as follows:

$$d(\mathrm{s}) = \mathbf{z}(\mathrm{s})^T \mathbf{z}(\mathrm{s}), \tag{8}$$

where $\mathbf{z}(\mathrm{s}) \triangleq \Sigma_0^{-1/2} (\mathbf{v}(\mathrm{s}) - \boldsymbol{\mu}_0)$. The random vector $\mathbf{z}(\mathrm{s})$ is distributed according to:

$$\begin{aligned} \mathbf{z}(\mathrm{s})|_{H_0} &\sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \\ \mathbf{z}(\mathrm{s})|_{H_1} &\sim \mathcal{N}\left(\Sigma_0^{-1/2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0), \Sigma_0^{-1}\Sigma_1\right). \end{aligned} \tag{9}$$

As such, the Mahalanobis distance under hypothesis H_0 is chi-square distributed with m degrees of freedom, regardless of the background clutter:

$$d(\mathrm{s})|_{H_0} \sim \chi_m^2(0). \tag{10}$$

The false alarm, as formulated in equation (2) is then given by:

$$P_{FA} = 1 - \mathrm{p}\left(\chi_m^2(0) \leqslant \eta\right). \tag{11}$$

The fact that the feature vector v(s) would be a Gaussian vector is surprising. Again in this paper the authors neglect the fact that they are doing multiple testing. They compute a probability of false alarms by reducing the test on a Gaussian test to a Chi2 test.

Defect detection in patterned wafers using anisotropic kernels

Maria Zontak · Israel Cohen

Algorithm 1 Defect Detection using NL-means estimation

{s - pixel index, f - source image, $\hat{f}$ - reconstructed source image}
for all $\mathbf{s} \in f$ **do**
  $P_{\mathbf{s}} \Leftarrow$ construct a patch of size $[s_x \times s_y]$ around pixel s
  $i \Leftarrow 1$
  {r - pixel index, f_{ref} - reference image, $\mathcal{N}_{\mathbf{s}}$- search region neighborhood of s}
  for all $\mathbf{r} \in \mathcal{N}_{\mathbf{s}}$ **do**
    $P_{\mathbf{r}}^{i} \Leftarrow$ construct a patch of size $[s_x \times s_y]$ around pixel r
    $\mathcal{W}^{i} \Leftarrow \exp(-\frac{\rho(P_{\mathbf{s}}, P_{\mathbf{r}}^{i})^2}{\epsilon})$ {ρ - a distance metric}
    $i \Leftarrow i + 1$
  $\mathcal{S}_{\mathcal{W}} \Leftarrow \Sigma_i \mathcal{W}^{i}$
  if $\mathcal{S}_{\mathcal{W}} = 0$ **then**
    for all i **do**
      $\mathcal{W}^{i} \Leftarrow 0$
  else
    for all i **do**
      $\mathcal{W}^{i} \Leftarrow \frac{\mathcal{W}^{i}}{\mathcal{S}_{\mathcal{W}}}$
  $\hat{P}_{\mathbf{s}} \Leftarrow \Sigma_{\forall i} \mathcal{W}^{i} \cdot P_{\mathbf{r}}^{i}$ {source image patch estimation using reference neighboring patches}
  $\mathcal{D}(\mathbf{s}) \Leftarrow \|\hat{P}_{\mathbf{s}} - P_{\mathbf{s}}\|_2$ {difference image value at pixel s calculation}
  $\hat{f}(\mathbf{s}) \Leftarrow \Sigma_{\forall i} \mathcal{W}^{i} \cdot f_{\text{ref}}(\mathbf{r}_i)$
  if $\hat{f}(\mathbf{s}) = 0$ **then**
    $\mathbf{s} \in \mathcal{A}$ {$\mathcal{A}$ is a set of defect regions}

This paper has similar aspects to what we are doing, but so confusing that I rubbed my eyes. The idea is to apply NL-means to all patches of the source image, NL-means being computed with respect to a reference image which is not itself anomalous. Here the NL-means parameter \$\epsilon\$ is crucial, because the threshold is on the sum of weights: if this sum is too small, then the patch is not reconstructed and detected as an anomaly. Thus \$\epsilon\$ is fixed just large enough so that any patch in the reference image can be reconstructed with non-zero weights. But if one thinks about it, one is led to the conclusion that the algorithm can be summarized much more simply as: a) fix a similarity threshold learned in the reference image and b) compare each patch of the source image to the patches of the reference; if the distance is higher than the similarity threshold, then the patch is an anomaly.

RARE2012: a multi-scale rarity-based saliency detection with its comparative statistical analysis

Rather than detailling the method, which is not stated in a very reproducible way, I describe the the principle. The idea is to build a saliency map based on rarity. To do so, at each point some 32 multiscale orientation features are computed using Gabor functions. But the most contrasted channels are privileged by a weight for reconstructing a unique orientation channel for each orientation. Then the histograms of these channels are computed and a pixel is given a weight which is roughty inversely proportional to its rarity in the histogram. The same idea is applied to the colors after PCA. Then summing all of these saliency maps one obtains something similar to what is observed with gaze trackers: the salient regions are the most visited. We could do the same directly by comparing a patch to all other patches and weighting inversely the patches that are less similar.

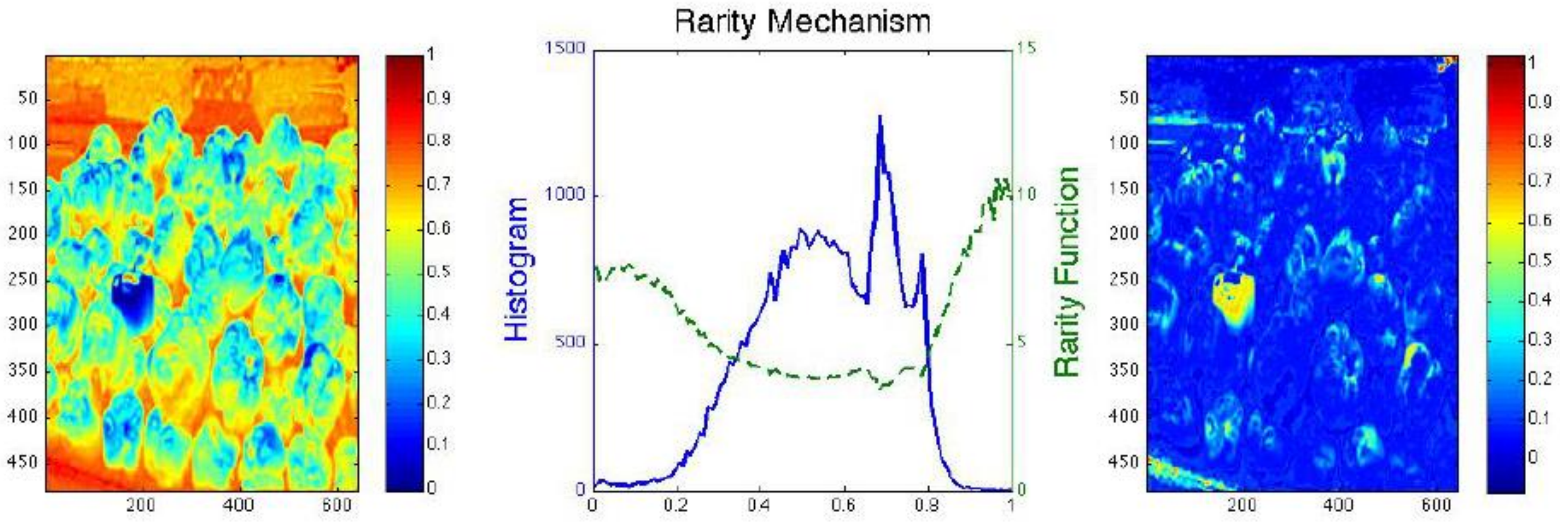


Figure 2: Illustration of the rarity mechanism on a single scale. Rarity function (green curve in the middle graph) is computed from a histogram (blue curve) of a feature map (left image) to a given scale. This process is repeated at several scales. Output is a reconstruction of the map where high values are given for the most "rare" areas (right image).

Exploiting Local and Global Patch Rarities for Saliency Detection Ali Borji Laurent Itti

We introduce a saliency model based on two key ideas. The first one is considering local and global image patch rarities as two complementary processes. The second one is based on our observation that for different images, oneof the RGB and Lab color spaces outperforms the other in saliency detection. We propose a framework that measures patch rarities in each color space and combines them in a final map. For each color channel, first, the input image is partitioned into non-overlapping patches and then each patch is represented by a vector of coefficients that linearly reconstruct it from a learned dictionary of patches from natural scenes. Next, two measures of saliency (Local and Global) are calculated and fused to indicate saliency of each patch. Local saliency is distinctiveness of a patch from its surrounding patches. Global saliency is the inverse of a patch's probability of happening over the entire image. The final saliency map is built by normalizing and fusing local and global saliency maps of all channels from both color systems.

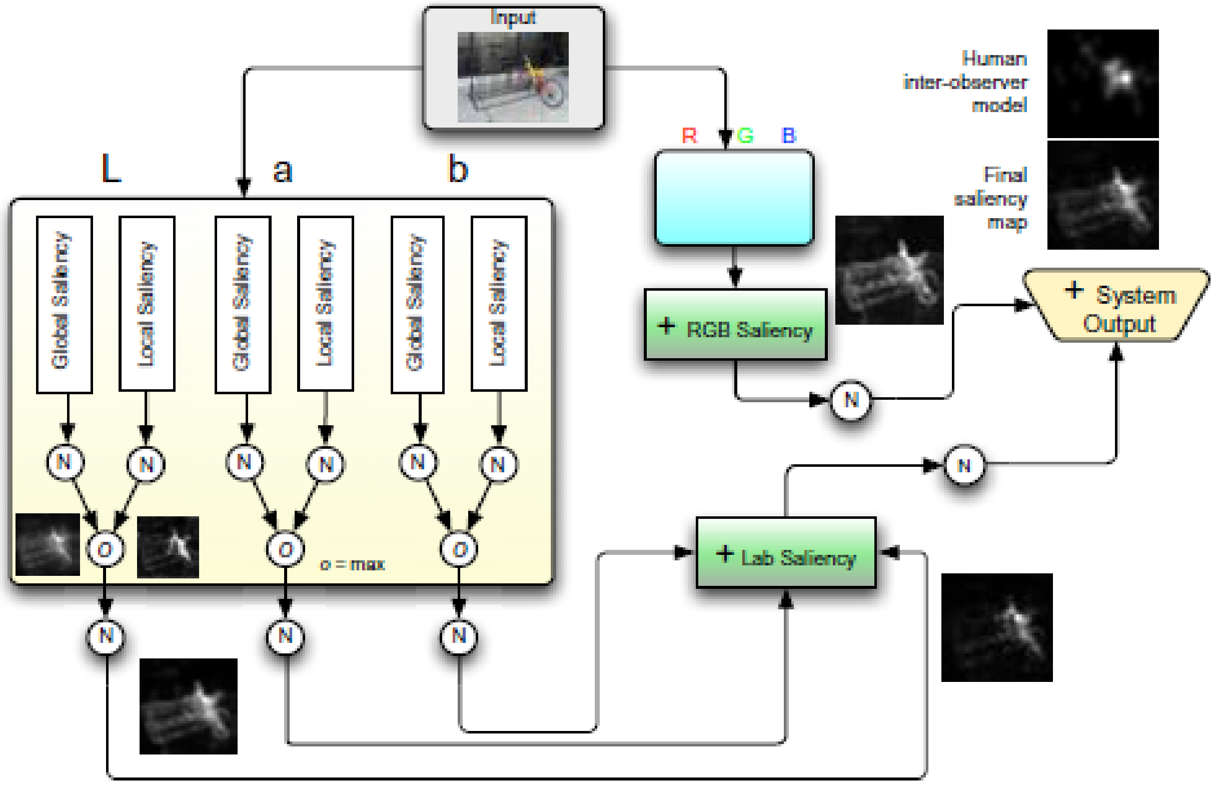


Figure 2. **Diagram of our proposed model.** First, the input image is transformed into Lab and RGB formats. Then, in each channel of a color space, a global saliency map based on rarity of an image patch in the entire scene, and a local saliency map, the dissimilarity between a patch and its surrounding window, are computed, normalized, and combined. Outputs of color channels (i.e., L, a, or b, similarly for RGB) are normalized and combined once more to form the output of a color system. The final map is the summation of the normalized maps in two color spaces.