KERNEL GENERALIZED GAUSSIAN AND ROBUST STATISTICAL LEARNING FOR ABNORMALITY DETECTION IN MEDICAL IMAGES

*Nitin Kumar, Ajit V. Rajwade, Sharat Chandran, Suyash P. Awate**

Computer Science and Engineering Department, Indian Institute of Technology (IIT) Bombay.

ABSTRACT

Typical methods for *abnormality detection* in medical images, which is a one-class classification problem, rely on kernel principal component analysis (KPCA) and its robust invariants. However, typical methods for robust KPCA appear heuristical in nature and often ignore the variances of the data along the principal modes of variation. In this paper, we propose a novel method for *robust* statistical learning in a reproducing kernel Hilbert space (RKHS) that relies on our extension of the *multivariate generalized Gaussian* distribution to *RKHS*. We propose novel algorithms to fit our kernel generalized Gaussian (KGG) in RKHS, using solely the Gram matrix and without the explicit lifting map. We exploit the KGG model, including mean, principal directions, and variances, for abnormality detection in medical images. The results on two large publicly available retinopathy datasets show that our method outperforms the state of the art.

Index Terms— Abnormality detection, one-class classification, kernel methods, robustness, generalized Gaussian.

1. INTRODUCTION AND RELATED WORK

Abnormality detection in medical image is a *one-class classification* problem [1], where training relies solely on data from the normal class. This is motivated by the difficulty of learning a model of abnormal image appearances stemming mainly from the tremendous variability in abnormalities. In this way, abnormality detection can be more challenging than a typical multi-class classification problem. Typical methods for abnormality detection, also called novelty detection, rely on principal component analysis (PCA) in input space or kernel PCA (KPCA) [2] in a kernel feature space.

It is well known that PCA and KPCA can be extremely sensitive to outliers in the data, leading to unreliable inference. In clinical applications of abnormality detection involving large training datasets intended to represent normal images, outliers can naturally arise because of specimen preparation (e.g., in microscopy), patient-related issues (e.g., motion), imaging artifacts, and manual mislabeling of abnormal images as normal. Although the literature presents several methods for increasing the *robustness* of KPCA-based approaches to outliers, the underlying algorithms are often heuristic in nature [3, 4, 5, 6]. For instance, [3, 4] employ adhoc rules for explicitly detecting outliers in the training set. While [7, 5, 6] describe robust KPCA (RKPCA) methods based on iterative data weighting schemes, it is unclear as to how to tune, or optimize, the weighting functions or their underlying parameters. One method in [7] first projects all data onto a sphere (unit norm), before performing KPCA, thereby distorting the original data. In contrast, we propose methods of RKPCA which address these issues through systematic statistical modeling and inference to provide robust estimations of the mean, principal directions of variation, and variances along the principal directions. During estimation, our method implicitly, and optimally, reweights the data, to reduce the effect of outliers, based on the estimated covariance structure of the data, unlike other methods that weight based on distance to mean. Typical methods [5, 6, 3, 4] compute robust means and modes of variation, but fail to compute and exploit variances along the principal modes. Thus, they perform poorly when the abnormal data lies within the subspace spanned by the normal data. In contrast, we propose a novel method to optimize, in addition to means and modes of variations, the associated variances to improve performance.

Some methods for robust PCA model learning [8, 9] rely on L_p norms ($p \geq 1$) in input space. In contrast, our method exploits L_q *quasi-norms* ($q > 0$) coupled with Mahalanobis distances in a reproducing kernel Hilbert space (RKHS). The method in [10] does *not* learn a PCA model from data, but proposes a robust projection of a corrupted datum onto a given PCA subspace, involving the difficult preimage problem.

Alternative popular kernel methods for abnormality detection rely on support vector machines (SVMs) [11], e.g., one-class SVM [12] and support vector data description (SVDD) [13]. Unlike KPCA, the SVM-based approaches are restricted to model a spherical distribution, or decision boundary, in RKHS and, hence, are known to be inferior to KPCA, both theoretically and empirically [2]. Furthermore, these SVM-based methods lack robustness to outliers in the training data. We propose methods that are improvements of KPCA leading to models that possess robustness to outliers in addition to the ability to model ellipsoidal or curved distributions (or decision boundaries) in RKHS.

*We are grateful for funding from Aditya Imaging Information Technologies and IIT Bombay Seed Grant 14IRCCSG010.

We characterize the class of normal images, or image regions, through their texture statistics using *textons* [14, 15] that capture the local geometric and photometric properties. Textons computed from image patches lead to models that can outperform models based on textons learned using filter banks [15, 16]. The texton-based model, for an image region, is a histogram of texton labels, indicating the (normalized) frequency of patches that are similar to the textons. Thus, we use kernel methods to model the nonlinear distribution of these histogram for normal images.

Typical methods for outlier detection in images [17, 18, 19, 20] rely on PCA, and often assume that the training set is very carefully selected and is outlier free. However, the outlier-free assumption is unrealistic. Typical methods like CHLOE [18] also rely on adhoc rules involving user-defined parameters to (i) eliminate outliers and (ii) weight data in training sets, based on, e.g., histograms and kurtosis of each individual feature in the feature set.

We propose a novel method for *robust kernel-based* statistical learning that extends the *multivariate generalized Gaussian* to RKHS, which we call the kernel generalized Gaussian (KGG). Our KGG model achieves robustness via L_q *quasi norms* on Mahalanobis distances in RKHS. We fit the KGG model using the Gram matrix only, without needing the explicit lifting map. We exploit the KGG model, including mean, principal directions, and *variances*, for abnormality detection in medical images. The results on two large publicly available retinopathy datasets show that the proposed method outperforms the state of the art.

2. KERNEL GENERALIZED GAUSSIAN (KGG)

We propose a novel robust statistical model relying on the multivariate generalized Gaussian in RKHS. In Euclidean space $\mathbb{R}^D$, the generalized Gaussian [21] is parametrized by: (i) mean vector $\mu \in \mathbb{R}^D$, (ii) covariance matrix $C \in \mathbb{R}^{D^2}$, and (iii) shape parameter $\rho \in \mathbb{R}_{\geq 0}$ (scalar). Varying ρ allows the generalized Gaussian to model a large class of multivariate statistical distributions, including Gaussian ($\rho = 2$), Laplacian ($\rho = 1$), and uniform ($\rho \to \infty$). We extend the generalized Gaussian to RKHS. We exploit $\rho < 1$, when the distribution has increased concentration near the mean and heavier tails, for robust fitting in the presence of outliers.

KGG Model. Consider a set of data points $\{x_n\}_{n=1}^N$ and a Mercer kernel $k(\cdot,\cdot)$ that implicitly maps the data to a RKHS $\mathcal{H}$ such that each datum x_n gets mapped to $\phi(x_n)$. Consider two vectors in RKHS: $f := \sum_{i=1}^I \alpha_i \phi(x_i)$ and $f' := \sum_{j=1}^J \beta_j \phi(x_j)$. The inner product $\langle f, f' \rangle_{\mathcal{H}} := \sum_{i=1}^I \sum_{j=1}^J \alpha_i \beta_j k(x_i, x_j)$. The norm $\|f\|_{\mathcal{H}} := \sqrt{\langle f, f \rangle_{\mathcal{H}}}$. When $f, f' \in \mathcal{H} \backslash \{0\}$, let $f \otimes f'$ be the rank-one operator defined as $f \otimes f'(g) := \langle f', g \rangle_{\mathcal{H}} f$. We fit a generalized Gaussian to the implicitly mapped data in RKHS, parametrized by mean vector $\mu \in \mathcal{H}$ and covariance operator C in the RKHS. The Mahalanobis distance of $f \in \mathcal{H}$ from mean $\mu \in \mathcal{H}$ relies on a regularized sample inverse-covariance operator $C^{-1} := \sum_{q=1}^Q (1/\lambda_q) v_q \otimes v_q$, where λ_q is the q-th largest eigenvalue of the covariance operator C, v_q is the corresponding eigenfunction, and $Q < N$ is a regularization parameter. We choose Q to be the number of principal eigenfunctions that capture 90% of the energy in the eigenspectrum. Then, the squared Mahalanobis distance of point $f \in \mathcal{H}$ from mean $\mu \in \mathcal{H}$ is $d^2_{\mathcal{M}}(f; \mu, C) := \langle f - \mu, C^{-1}(f - \mu) \rangle_{\mathcal{H}}$. We define a generalized Gaussian in RKHS as

$$P(f) := \frac{\delta(\rho/2)}{2|C|^{0.5}} \exp\left[-\Big(\eta(\rho/2) d^2_{\mathcal{M}}(f; \mu, C) \Big)^{\rho/2} \right], \quad (1)$$

where $\delta(r) := r\Gamma(2/r)/(\pi\Gamma(1/r)^2)$ is a normalization constant, $\eta(r) := \Gamma(2/r)/(2\Gamma(1/r))$, and $|C| := \prod_{q=1}^Q \lambda_q$.

We fit the KGG model to the implicitly mapped data in RKHS by (i) first estimating the robust mean μ, (ii) then estimating the robust principal directions $\{v_q\}_{q=1}^Q$, and followed by (iii) estimating the robust variances $\{\lambda_q\}_{q=1}^Q$. In our model, ρ is a free parameter (fixed during parameter estimation) that we tune based on the application, using training data. We show that the model fitting requires solely the Gram matrix K, where $K_{ij} := \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}} = k(x_i, x_j)$, without requiring to know the mapping function $\phi(\cdot)$.

KGG Model Fitting: Estimating Mean. The mean μ must be in the span of the mapped data $\{\phi(x_i)\}_{i=1}^N$. Thus, we represent the mean, using a weight vector $\beta \in \mathbb{R}^N$, as $\mu(\beta) := \sum_{i=1}^N \beta_i \phi(x_i)$. Estimating μ is then equivalent to estimating β. We define the optimal weights $\beta^* := \arg\min_\beta f(\beta)$, where

$$f(\beta) := \sum_{i=1}^N (\|\phi(x_i) - \mu(\beta)\|_{\mathcal{H}})^\rho. \quad (2)$$

We rewrite the objective function $f(\beta)$ as

$$f(\beta) = \sum_{i=1}^N \left(\beta^\top K \beta + K_{ii} - 2\beta^\top K_i \right)^{\rho/2}, \quad (3)$$

where K_i is the i-th column of K. We optimize via iterative gradient descent with adaptive step size (adjusted at each update) to ensure that each update reduces the objective function. We initialize mean μ to the sample mean: $\beta_i := 1/N$.

When $\rho = 2$, the estimated mean is the sample mean that is affected by outliers. As ρ reduces from $2 \to 0$, the effect of the outliers decreases in the objective function. Correspondingly, the gradient term for an outlier j is weighted down far more (with large $\|\phi(x_j) - \mu(\beta)\|_{\mathcal{H}}$) than for the inliers. This leads to a mean estimate that is robust to outliers.

KGG Model Fitting: Estimating Eigenvectors. Given the estimated mean $\mu(\beta^*)$, we estimate the first principal direction v_1 by maximizing the ρ-th moment (around the mean) of the projected data onto the direction v_1. For $\rho = 2$, this is equivalent to maximizing the variance. Let the centered data in RKHS be $\widetilde{\phi}(x_i) := \phi(x_i) - \mu(\beta^*)$. Because each eigenvector must lie in the span of the data, we represent the m-th eigenvector, using the weight vector $\alpha_m \in \mathbb{R}^N$, as $v(\alpha_m) := \sum_{j=1}^{N} \alpha_{mj} \widetilde{\phi}(x_j)$. After estimating the first $m-1$ principal directions, i.e., $v_1, \cdots, v_{m-1}$, we define the optimal weights $\alpha_m^* := \arg\max_{\alpha_m} g(\alpha_m)$, where

$$g(\alpha_m) := \sum_{i=1}^{N} \left[(\langle \widetilde{\phi}(x_i), v(\alpha_m) \rangle_{\mathcal{H}})^2 \right]^{\rho/2}. \quad (4)$$

and $v(\alpha_m)$ is constrained to (i) unit norm, i.e., $\|v(\alpha_m)\|_{\mathcal{H}} = 1$, and (ii) be orthogonal to all previously estimated principal directions, i.e., $\langle v(\alpha_m), v_l \rangle_{\mathcal{H}} = 0, \forall 1 \leq l \leq m-1$. Let the centered-data Gram matrix be $\widetilde{K}$, where $\widetilde{K}_{ij} := \langle \widetilde{\phi}(x_i), \widetilde{\phi}(x_j) \rangle_{\mathcal{H}}$. Then, we rewrite the objective function as

$$g(\alpha_m) = \sum_{i=1}^{N} \left[(\widetilde{K}_i^\top \alpha_m)^2 \right]^{\rho/2}. \quad (5)$$

and rewrite the constraints as (i) $\alpha_m^\top \widetilde{K} \alpha_m = 1$ (unit norm) and (ii) $\alpha_l^\top \widetilde{K} \alpha_m = 0, \forall 1 \leq l \leq m-1$ (orthogonality). We optimize using projected gradient descent with adaptive step size. All orthogonality constraints are linear in α_m, implying projections of α_m onto associated hyperplanes in $\mathbb{R}^N$. The projection required for the unit norm constraint is performed via [22] that deals with projecting point on hyper-ellipsoids; in practice, we find that simply scaling α_m by $(\alpha_m^\top \widetilde{K} \alpha_m)^{0.5}$ works sufficiently well if we take small gradient steps, keeping successive α_m close to the constraint set. We initialize v_m to the optimal solution for $\rho = 2$ that is equivalent to KPCA.

KGG Model Fitting: Estimating Variances λ_q. Given the estimated mean $\mu(\beta^*)$ and the estimated eigenvectors $\{v_q(\alpha_q^*)\}_{q=1}^{Q}$, we optimize the variances $\Lambda := \{\lambda_q\}_{q=1}^{Q}$ along the modes of variation $\{v_q\}_{q=1}^{Q}$ by minimizing the negative log likelihood of the implicitly mapped data in RKHS. Thus, we define the optimal variance $\Lambda^* := \arg\min_{\Lambda} h(\Lambda)$, where

$$h(\Lambda) := \sum_{i=1}^{N} \left[\frac{\log(\lambda_q)}{2} + \sum_{q=1}^{Q} \left(\frac{\langle \widetilde{\phi}(x_i), v_q \rangle_{\mathcal{H}}^2}{\lambda_q} \right)^{\rho/2} \right], \quad (6)$$

under the constraint that $\lambda_q > 0, \forall 1 \leq q \leq Q$. When $v(\alpha_m) := \sum_{j=1}^{N} \alpha_{mj} \widetilde{\phi}(x_j)$, the inner product $\langle \widetilde{\phi}(x_i), v(\alpha_m) \rangle_{\mathcal{H}}$ is easily evaluated as $\alpha_m^\top \widetilde{K}_i$.

KGG for Abnormality Detection. After fitting our KGG model to data, we define a decision boundary enclosing the normal class through a threshold on the Mahalanobis distance computed from the mean such that 98.5% of the probability mass lies within the boundary. This threshold varies with the value of ρ; for the standard Gaussian case with $\rho = 2$ and the univariate setting with variance σ^2, the threshold corresponds to the distance of $\pm 2.5\sigma$. For the univariate generalized Gaussian case, this threshold depends on ρ and can be computed in a straightforward manner from the inverse cumulative distribution function that is known analytically. Knowing that the threshold is based on the Mahalanobis distance that is independent of the scale parameter, the threshold naturally extends to the multivariate case.

3. RESULTS AND DISCUSSION

We evaluate our method for abnormality detection on simulated data and real-world image data. We evaluate all methods on large publicly available datasets that provide a large number of examples of the normal and abnormal classes both. Indeed, the training, i.e., model learning, for abnormality detection methods relies solely on data from the normal class, which includes outliers and mislabelled data incorrectly labelled to the normal class. We compare our KGG method with 5 other methods: (i) standard KPCA [23], which is a special case of KGG when $\rho = 2$, (ii) Huang et al.'s RKPCA [6], (iii) one-class regularized kernel SVM [12], (iv) regularized kernel SVDD [13], and (v) two-class regularized kernel SVM [11] that, unlike the aforementioned methods, relies on training data. We use cross validation to tune the free parameters underlying all methods, i.e., the kernel parameters, ρ (for our method), and the regularization parameters for SVM-based methods.

3.1. Results on Simulated Data

We simulate data in 2D Euclidean space to mimic what a real-world dataset would lead to in RKHS (after kernel-based mapping). We simulate data (Figure 1) from a Gaussian (normal class): mean $(0, 5)$, modes of variation as the cardinal axes, and standard deviations along the modes of variation as $(0.5, 2)$. We then contaminate the data with outliers of two kinds: (i) spread uniformly over the domain; (ii) clustered at a location far away. For training, the normal-class sample size is 5000 contaminated with 1000 outliers. For testing, the normal-class sample size is 5000 and abnormal-class sample size is 3000. We choose the kernel similarity as the Euclidean inner-product. Our KGG model learning is far more robust to outliers in the training data with a classification accuracy of 93%, outperforming (i) PCA (accuracy 69%), (ii) SVDD (accuracy 61%), and (iii) 2-class SVM (accuracy 71%).

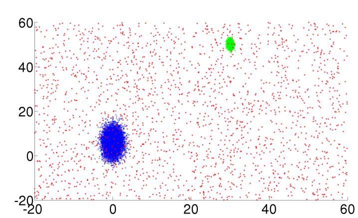

Fig. 1. **Results on Simulated Data.** Data from a 2D Gaussian distribution (blue) contaminated with outliers (red and green).

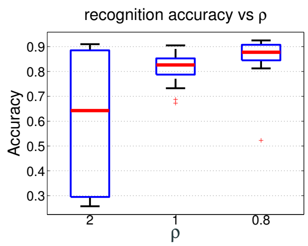

Fig. 2. **Results on MNIST Data.** Recognition accuracy with our method as a function of the shape parameter ρ. The box plots show variability in recognition performance for resampling of the training data (30 repeats).

3.2. Results on Handwritten Digit Image Recognition

We take the MNIST dataset [24] and consider (i) images of digit 0 as the "normal" class and (ii) images of digit $1-9$ as "abnormal". We take the training data as the images of digit 0. It is well known [12] that this dataset has outliers inherently as part of the dataset, arising from poor handwriting and mislabeling. We consider the image itself as the feature. The recognition accuracy for our KGG method (Gaussian kernel) increases as ρ decreases from 2 (KPCA) to 0.8 (Figure 2). This shows the utility of: (i) robust learning for real-world datasets and (ii) our method exploiting quasi norms when $\rho < 1$.

3.3. Results on Real-World Medical Image Datasets

We use 2 large publicly available datasets of retinopathy images: Messidor [25] (Figure 3) and Kaggle [26] (Figure 4). The figures show that both datasets, even though carefully constructed, already have outliers in the normal class. We use the texton-based histogram feature, using patches (9×9) to compute textons, to classify regions (50×50) as normal or abnormal. We use the intersection kernel [27]. From the datasets, we select a training set with 12000 normal images and, to mimic a real-world scenario, contaminate it by adding another 3% of abnormal images mislabeled as normal. We select a test set having 8000 normal images and 5000 abnormal images. The accuracy of abnormality detection for our KGG approach outperforms all other methods for both Messidor and Kaggle datasets (Figure 5).

Conclusion. We have proposed a novel method for *robust kernel-based* statistical learning that relies on our generalization of the *multivariate generalized Gaussian* to RKHS. We propose novel algorithms to fit our KGG in RKHS, using solely the Gram matrix and without the explicit lifting map. We exploit the KGG model, including its *covariance* operator, for abnormality detection in real-world medical applications where a small fraction of training data is inevitably contaminated because of outliers and mislabeling. The results on two large publicly available retinopathy datasets show that the proposed method outperforms the state of the art, including one-class classification methods (KPCA, one-class kernel SVM, kernel SVDD) and the two-class kernel SVM.

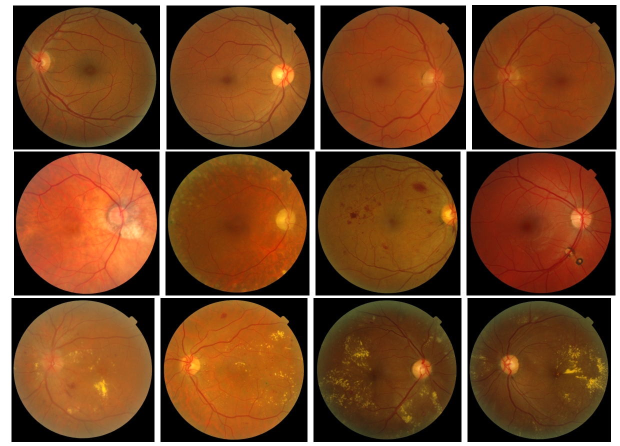

Fig. 3. **Retinopathy Dataset–Messidor. Top Row:** Normal images. **Middle Row:** Images labelled normal, but are outliers. **Bottom Row:** Abnormal images.

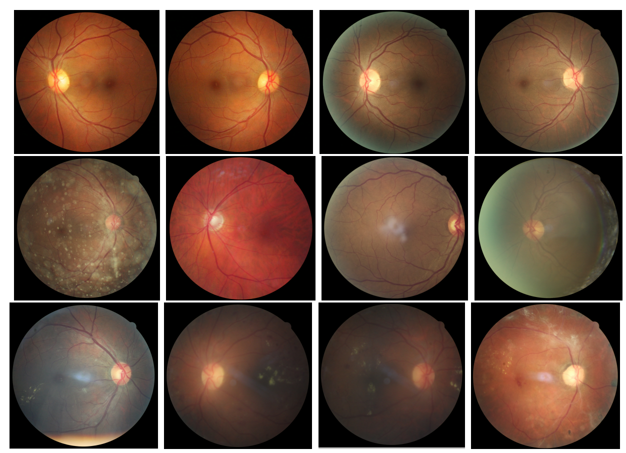

Fig. 4. **Retinopathy Dataset–Kaggle. Top Row:** Normal images. **Middle Row:** Images labelled normal, but are outliers. **Bottom Row:** Abnormal images.

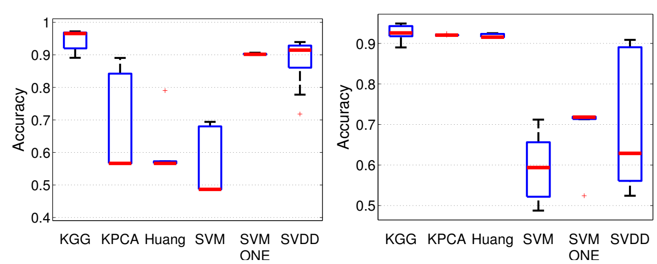

Fig. 5. **Results on Retinopathy Datasets.** Classification accuracy using learning on training sets contaminated with outliers, for: **(Left)** Messidor dataset; **(Right)** Kaggle dataset. The box plots show variability in recognition performance for resampling of the training data (20 repeats).

4. REFERENCES

[1] M Moya and D Hush, "Network constraints and multi-objective optimization for one-class classification," *Neural Net.*, vol. 9, no. 3, pp. 463–74, 1996.

[2] H Hoffmann, "Kernel PCA for novelty detection," *Pattern Recog.*, vol. 40, no. 3, pp. 863–874, 2007.

[3] Y Li, "On incremental and robust subspace learning," *Pattern Recog.*, vol. 37, no. 7, pp. 1509–18, 2004.

[4] C Lu, T Zhang, R Zhang, and C Zhang, "Adaptive robust kernel PCA algorithm," in *IEEE Int. Conf. Acoustics, Speech, and Sig. Proc.*, 2003, vol. 6, p. 621.

[5] S Huang, Y Yeh, and S Eguchi, "Robust kernel principal component analysis," *Neural Comp.*, vol. 21, no. 11, pp. 3179–213, 2009.

[6] H Huang and Y Yeh, "An iterative algorithm for robust kernel principal component analysis," *Neurocomputing*, vol. 74, no. 18, pp. 3921–30, 2011.

[7] M Debruyne and T Verdonck, "Robust kernel principal component analysis and classification," *Adv. Data Anal. Classif.*, vol. 4, no. 2, pp. 15167, 2010.

[8] N Kwak, "Principal component analysis based on L1-norm maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 9, pp. 1672–80, 2008.

[9] N Kwak, "Principal component analysis by Lp-norm maximization," *IEEE Trans. Cybern.*, vol. 44, no. 5, pp. 594–609, 2014.

[10] M Nguyen and F Torre, "Robust kernel principal component analysis," in *Adv. Neural Info. Proc. Sys.*, 2008, pp. 1185–92.

[11] C Cortes and V Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–97, 1995.

[12] B Scholkopf, J Platt, J Shawe-Taylor, A Smola, and R Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comp.*, vol. 13, no. 7, pp. 1443–71, 2001.

[13] D Tax and R Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, 2004.

[14] T Leung and J Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. J. Comp. Vis.*, vol. 43, no. 1, pp. 29–44, 2001.

[15] M Varma and A Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2032–47, 2009.

[16] L Wolf, X Huang, I Martin, and D Metaxas, "Patch-based texture edges and segmentation," *Proc. Euro. Conf. Comp. Vis.*, pp. 481–93, 2006.

[17] R Norousi, S Wickles, C Leidig, T Becker, V Schmid, R Beckmann, and A Tresch, "Automatic post-picking using MAPPOS improves particle image detection from cryo-em micrographs," *J. Struct. Biol.*, vol. 182, no. 2, pp. 59–66, 2013.

[18] S Manning and L Shamir, "CHLOE: A software tool for automatic novelty detection in microscopy image datasets," *J. Open Res. Soft.*, vol. 2, no. 1, 2014.

[19] J Vargas, V Abrishami, R Marabini, J Rosa-Trevin, A Zaldivar, J Carazo, and C Sorzano, "Particle quality assessment and sorting for automatic and semiautomatic particle-picking techniques," *J. Struct. Biol.*, vol. 183, no. 3, pp. 342–53, 2013.

[20] C Sorzano, J Vargas, J Rosa-Trevin, A Zaldivar-Peraza, J Oton, V Abrishami, I Foche, R Marabini, G Caffarena, and J Caraza, "Outlier detection for single particle analysis in electron microscopy," in *Int. Work-Conf. Bioinfo. Biomed. Engg.*, 2014, pp. 950–9.

[21] M Novey, T Adali, and A Roy, "A complex generalized Gaussian distribution - characterization, generation, and estimation," *IEEE Trans. Signal Proc.*, vol. 58, no. 3, pp. 1427–33, 2010.

[22] Y Dai, "Fast algorithms for projection on an ellipsoid," *SIAM J. Optim.*, vol. 16, no. 4, pp. 986–1006, 2006.

[23] B Schlkopf, A Smola, and K Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comp.*, vol. 10, no. 5, pp. 1299–1319, 1998.

[24] Y LeCun, L Bottou, Y Bengio, and P Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–324, 1998.

[25] E Decenciere, X Zhang, G Cazuguel, B Lay, B Cochener, C Trone, P Gain, R Ordonez, P Massin, and A Erginay, "Feedback on a publicly distributed image database: the Messidor database," *Imag. Anal. Stereology*, vol. 33, no. 3, pp. 231–4, 2014.

[26] "Diabetic retinopathy data," `www.kaggle.com/c/diabetic-retinopathy-detection/data`.

[27] A Barla, F Odone, and A Verri, "Histogram intersection kernel for image classification," in *IEEE Int. Conf. Imag. Proc.*, 2003, vol. 3, p. 513.