Kernel Generalized-Gaussian Mixture Model for Robust Abnormality Detection

Nitin Kumar, Ajit V. Rajwade, Sharat Chandran, and Suyash P. Awate(✉)

Computer Science and Engineering Department,
Indian Institute of Technology (IIT) Bombay, Mumbai, India
suyash@cse.iitb.ac.in

Abstract. Typical methods for *abnormality detection* in medical images rely on principal component analysis (PCA), kernel PCA (KPCA), or their robust invariants. However, typical robust-KPCA methods use heuristics for model fitting and perform outlier detection ignoring the variances of the data within principal subspaces. In this paper, we propose a novel method for *robust* statistical learning by extending the *multivariate generalized-Gaussian* distribution to a *reproducing kernel Hilbert space* and employing it within a *mixture model*. We propose *expectation maximization* to fit our kernel generalized-Gaussian mixture model (KGGMM), using solely the Gram matrix and without the explicit lifting map. We exploit the KGGMM, including component means, principal directions, and variances, for abnormality detection in images. The results on 4 large publicly available datasets, involving retinopathy and cancer, show that our method outperforms the state of the art.

Keywords: Abnormality detection · One-class classification · Kernel methods · Robustness · Generalized gaussian · Mixture model · Expectation maximization

1 Introduction and Related Work

Abnormality detection in medical images [3,10] is a *one-class classification* problem [13], where training relies solely on data from the normal class. This is motivated by the difficulty of learning a model of abnormal image appearances because of their tremendous variability. Typical methods for abnormality detection rely on principal component analysis (PCA) or kernel PCA (KPCA) [4].

In clinical applications involving large training datasets intended to represent normal images, *outliers* naturally arise because of errors in specimen preparation (e.g., slicing or staining in microscopy), patient issues (e.g., motion), imaging artifacts, and manual mislabeling of abnormal images as normal. KPCA is very sensitive to outliers in the data, leading to unreliable inference. Some methods for abnormality detection [11] rely on PCA, assuming training sets to be outlier free. Typical *robust* KPCA (RKPCA) methods [3,5,7–9] are heuristic in their

The authors are grateful for funding from Aditya Imaging Information Technologies. S.P. Awate thanks funding from IIT Bombay Seed Grant 14IRCCSG010.

M. Descoteaux et al. (Eds.): MICCAI 2017, Part III, LNCS 10435, pp. 21–29, 2017.
DOI: 10.1007/978-3-319-66179-7_3

modeling and inference. For instance, [7–9] employ adhoc rules for explicitly detecting outliers in the training set. While [2,5] describe RKPCA based on iterative data-weighting, using distance to the mean, the weighting functions seem adhoc. CHLOE [9] also uses rules involving free parameters to weight data based on kurtosis of individual features. One method [2] distorts data by projecting it onto a sphere (unit norm). In contrast, we propose a method using statistical (mixture) modeling to infer robust estimates of means and covariances. During estimation, our method implicitly, and optimally, reweights the data, to reduce the effect of outliers, based on the covariance structure of the data.

Typical abnormality detection methods [3,5,7,8] compute robust means and modes of variation, but fail to compute and exploit variances along the modes. Thus, they perform poorly when the abnormal data lies within the subspace spanned by the normal data. In contrast, our method optimizes, in addition to means and modes, the associated variances to improve performance. Some methods [3,6] for robust PCA model learning 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).

Some kernel methods for abnormality detection rely on the support vector machine (SVM), e.g., one-class SVM [13] and support vector data description (SVDD) [15]. Unlike KPCA, these SVM methods model only a spherical distribution or decision boundary in RKHS and, thus, are inferior to KPCA theoretically and empirically [4]. Also, the SVM methods lack robustness to outliers in the training data. In contrast, our method is robust to outliers and enables us to model arbitrarily curved distributions as well as decision boundaries in RKHS.

We propose a novel method for *robust* statistical learning by extending the *multivariate generalized-Gaussian* distribution to a *RKHS* for *mixture modeling*. We propose *expectation maximization* (EM) to fit our kernel generalized-Gaussian mixture model (KGGMM), using solely the Gram matrix, without the explicit lifting map. We model geometric and photometric properties of image texture via standard texton-label histograms [16]. We exploit the KGGMM, including component means, principal directions, and *variances*, for abnormality detection. The results on 4 large publicly available datasets, involving retinopathy and cancer, show that our method outperforms the state of the art.

2 Methods

In $\mathbb{R}^D$, the generalized Gaussian [12] is parametrized by the mean $\mu \in \mathbb{R}^D$, covariance matrix $C \in \mathbb{R}^{D \times D}$, and shape $\rho \in \mathbb{R}_{>0}$; Gaussian ($\rho = 2$), Laplacian ($\rho = 1$), uniform ($\rho \to \infty$). We extend the generalized Gaussian to RKHS for mixture modeling. We exploit $\rho < 1$, when the distribution has increased concentration near the mean and heavier tails, for robust fitting amidst outliers.

2.1 Kernel Generalized Gaussian (KGG)

Consider a set of N data points $\{x_n \in \mathbb{R}^D\}_{n=1}^N$ in input space. Consider a Mercer kernel $\kappa(\cdot, \cdot)$ that implicitly maps the data to a RKHS $\mathcal{H}$ such that each datum

x_n gets mapped to $\phi(x_n)$. Consider 2 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 $\langle f, f' \rangle_{\mathcal{H}} := \sum_{i=1}^{I} \sum_{j=1}^{J} \alpha_i \beta_j \kappa(x_i, x_j)$. 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 $f \otimes f'(g) := \langle f', g \rangle_{\mathcal{H}} f$. The generalized Gaussian extended to RKHS is parametrized by shape $\rho \in \mathbb{R}_{>0}$, mean $\mu \in \mathcal{H}$, and covariance operator $C = \sum_{q=1}^{Q} \lambda_q v_q \otimes v_q$, where λ_q is the q-th largest eigenvalue of covariance C, v_q is the corresponding eigenfunction, and $Q < N$ is a regularization parameter. We set Q to the number of principal eigenfunctions that capture 95% of the eigenspectrum energy. For $f \in \mathcal{H}$, the squared Mahalanobis distance is $d^2_{\mathcal{M}}(f; \mu, C) := \langle f - \mu, C^{-1}(f - \mu) \rangle_{\mathcal{H}}$, where $C^{-1} = \sum_{q=1}^{Q} (1/\lambda_q) v_q \otimes v_q$ is the sample inverse-covariance operator. Then, our generalized Gaussian in RKHS is

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

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

2.2 Kernel Generalized-Gaussian Mixture Model (KGGMM)

We propose to model the distribution of data $x := \{x_n \in \mathbb{R}^D\}_{n=1}^{N}$ using a Mercer kernel to implicitly map the data to a RKHS, i.e., $\{\phi(x_n) \in \mathcal{H}\}_{n=1}^{N}$, and then representing the distribution in RKHS using a mixture of KGG distributions. Consider a KGG mixture model with K components, where the k-th component is the KGG $P_{\mathcal{G}}(\cdot; \mu_k, C_k, \rho)$ coupled with weight $\omega_k \in \mathbb{R}_{\geq 0}$, such that $\omega_k \leq 1$ and $\sum_{k=1}^{K} \omega_k := 1$. For each datum x_n, let Z_n be the *hidden* (label) random variable indicating the mixture component from which the datum was drawn.

Each mean μ_k must lie in the span of the mapped data $\{\phi(x_i)\}_{i=1}^{N}$. Thus, we represent each mean, using coefficient vector $\beta_k \in \mathbb{R}^N$, as $\mu_k(\beta_k) := \sum_{i=1}^{N} \beta_{ki} \phi(x_i)$. Estimating μ_k is then equivalent to estimating β_k. We represent each covariance operator C_k using its Q principal eigenvectors $\{v_{kq} \in \mathcal{H}\}_{q=1}^{Q}$ and eigenvalues $\{\lambda_{kq} \in \mathbb{R}_{>0}\}_{q=1}^{Q}$. Each eigenvector of C_k must lie in the span of the mapped data. So, we represent the q-th eigenvector of C_k, using coefficient vector $\alpha_{kq} \in \mathbb{R}^N$, as $v_{kq}(\alpha_{kq}) := \sum_{j=1}^{N} \alpha_{kqj} \phi(x_j)$. Estimating v_{kq} is equivalent to estimating α_{kq}.

Model Fitting. We propose EM to fit the KGGMM to the mapped data to maximize the likelihood function. The prior label probability $P(z_n = k) := \omega_k$. The complete-data likelihood $P(z, x) := \prod_{n=1}^{N} P(z_n) P_{\mathcal{G}}(\phi(x_n); \mu_{z_n}, C_{z_n}, \rho)$. We show that EM does *not* need the map $\phi(\cdot)$, but only the Gram matrix G, where $G_{ij} := \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}} = \kappa(x_i, x_j)$. In our framework, ρ is a free parameter (fixed before EM) that we tune using training data; $\rho < 1$ gives best results.

Initialization. We use kernel k-means to initialize the parameters. We initialize (i) mean μ_k to the k-th cluster center, (ii) weight ω_k to the fraction of data assigned to cluster k, and (iii) covariance C_k using KPCA on cluster k.

E Step. At the t-th iteration, let the set of parameters be $\theta^t := \{\beta_k^t \in \mathbb{R}^N, \{\alpha_{kq}^t \in \mathbb{R}^N\}_{q=1}^Q, \{\lambda_{kq}^t \in \mathbb{R}\}_{q=1}^Q, \omega_k^t \in \mathbb{R}\}_{k=1}^K$. Let α_k denote a $N \times Q$ matrix, representing the Q eigenfunctions of C_k, such that its q-th column is α_{kq}. Let λ_k denote a $Q \times Q$ diagonal matrix, representing the Q eigenvalues of C_k, such that its q-th diagonal element is λ_{kq}. Given θ^t, the E step defines the function $Q(\theta; \theta^t) := E_{P(Z|x,\theta^t)}[\log P(Z, x; \theta)]$ that can be simplified to

$$\sum_{n=1}^{N}\sum_{k=1}^{K}\gamma_{nk}^t\left[\log\omega_k - \sum_{q=1}^{Q}\left[\frac{\log(\lambda_{kq})}{2} + \left(\eta\left(\frac{\rho}{2}\right)\frac{\langle\phi(x_n) - \mu_k(\beta_k), v_{kq}(\alpha_{kq})\rangle_{\mathcal{H}}^2}{\lambda_{kq}}\right)^{\frac{\rho}{2}}\right]\right]$$

excluding terms independent of θ, and where the membership of datum x_n to mixture component k, given the current parameter estimate θ^t, is the posterior $\gamma_{nk}^t := P(Z_n = k|x_n, \theta^t) = \omega_k^t P_{\mathcal{G}}(\phi(x_n); \mu_k, C_k, \rho)/P(x_n; \theta^t)$ by Bayes rule.

M Step. The M step updates parameter estimates to $\theta^{t+1} := \arg\max_\theta Q(\theta; \theta^t)$ subject to constraints on: (i) weights, such that $\omega_k \geq 0, \sum_k \omega_k = 1$, (ii) eigenvalues, such that $\lambda_{kq} > 0$, and (iii) coefficients, such that eigenvectors $v_{kq}(\alpha_{kq})$ are unit norm ($\|v_{kq}\|_{\mathcal{H}} = 1$) and mutually orthogonal ($\langle v_{kq}, v_{kr}\rangle_{\mathcal{H}} = 0, \forall q \neq r$).

Estimating Weights. The optimal weights ω_k^{t+1} are given by the solution to $\arg\max_\omega \sum_{n=1}^N \sum_{k=1}^K \gamma_{nk}^t \log\omega_k$, subject to the positivity and sum-to-unity constraints. The method of Lagrange multipliers gives $\omega_k^{t+1} = \sum_{n=1}^N \gamma_{nk}^t/N$.

Estimating Means. Given weights ω_k^{t+1}, the optimal mean $\mu_k^{t+1}(\beta_k^{t+1})$ is given by $\beta_k^{t+1} := \arg\min_{\beta_k} \sum_n \gamma_{nk}^t \sum_q \left((G_n^\top \alpha_{kq} - \beta_k^\top G\alpha_{kq})^2/\lambda_{kq}\right)^{\rho/2}$, where G_n is the n-th column of the Gram matrix G. We optimize via gradient descent with adaptive step size (adjusted at each update) to ensure that each update improves the objective function value. When $\rho = 2$, the mean estimate is the (weighted) sample mean that is affected by outliers. As ρ reduces, the effect of the outliers decreases in the objective function; the gradient term for an outlier j is weighted down far more than for the inliers, leading to robust estimates.

Estimating Eigenvectors. Given weights ω_k^{t+1} and means $\mu_k^{t+1}(\beta_k^{t+1})$, the optimal set of eigenfunctions $v_k^{t+1}(\alpha_k^{t+1})$ is given by $\alpha_k^{t+1} := \arg\min_{\alpha_k} \sum_n \gamma_{nk}^t \left[(\alpha_k^\top G_n - \alpha_k^\top G\beta_k^{t+1})^\top \lambda_k^{-1} (\alpha_k^\top G_n - \alpha_k^\top G\beta_k^{t+1})\right]^{\rho/2}$, subject to orthonormality constraints on the set of eigenfunctions $\{v_{kq}(\alpha_{kq})\}_{q=1}^Q$. We optimize via projected gradient descent with adaptive step size, where each step (i) first uses a gradient-descent step to update matrix α_k to $\widetilde{\alpha}_k$, implicitly updating the eigenfunctions to $\{\widetilde{v}_{kq}(\widetilde{\alpha}_{kq})\}_{q=1}^Q$, and (ii) then updates $\widetilde{\alpha}_k$ to α_k^{t+1} by projecting the eigenfunction set $\{\widetilde{v}_{kq}(\widetilde{\alpha}_{kq})\}_{q=1}^Q$ onto the space of orthogonal eigenfunction bases. In Euclidean space, the projection of a set of Q vectors, represented as the columns of a matrix M, onto the space of Q orthogonal vectors is given by $LR^\top$ where matrices L and R comprise the left and right singular vectors in the singular value decomposition (SVD) of M. In Euclidean space, $LR^\top = M(M^\top M)^{-0.5}$. In a RKHS, we replace the SVD by the kernel SVD as follows.

Consider Q functions $F := \{f_q \in \mathcal{H}\}_{q=1}^{Q}$ that are *not* orthogonal. Let the kernel SVD of F be the operator $\sum_{q=1}^{Q} s_q a_q \otimes b_q$, where the singular values are $s_q \in \mathbb{R}_{\geq 0}$, and the left and right singular vectors are the orthonormal sets $\{a_q \in \mathcal{H}\}_{q=1}^{Q}$ and $\{b_q \in \mathbb{R}^Q\}_{q=1}^{Q}$, respectively. Consider the $Q \times Q$ matrix Y where $Y_{ij} := \langle f_i, f_j \rangle_{\mathcal{H}}$. The matrix Y also equals $\sum_{q'=1}^{Q} s_{q'} b_{q'} \otimes a_{q'} (\sum_{q''=1}^{Q} s_{q''} a_{q''} \otimes b_{q''})$ that reduces to $\sum_{q=1}^{Q} s_q^2 b_q b_q^\top$ because of the orthogonality of the left singular vectors. Thus, an eigen decomposition of the matrix Y yields the eigenvalues as s_q^2 and the eigenvectors as b_q. Subsequently, we observe that the required projection of F onto the space of orthogonal functions in RKHS is given by $\sum_{q=1}^{Q} s_q a_q \otimes b_q (Y^{-0.5}) = \sum_{q=1}^{Q} s_q a_q \otimes b_q (\sum_{q'=1}^{Q} s_{q'}^{-1} b_{q'} b_{q'}^\top) = \sum_{q=1}^{Q} a_q \otimes b_q$. In practice, when we represent the eigenvectors using the $N \times Q$ matrix $\widetilde{\alpha}_k$, the matrix $Y = \widetilde{\alpha}_k^\top G \widetilde{\alpha}_k$ and the projection gives us $\alpha_k^{t+1} = \widetilde{\alpha}_k (Y)^{-0.5}$.

Estimating Variances. Given weights ω_k^{t+1}, means $\mu_k^{t+1}(\beta_k^{t+1})$, and eigenfunctions $v_k^{t+1}(\alpha_k^{t+1})$, each optimal eigenvalue is given by $\lambda_{kq}^{t+1} := \arg\min_{\lambda_{kq}>0} \sum_{n=1}^{N} \gamma_{nk}^t [0.5 \log(\lambda_{kq}) + (\eta(\rho/2) a_{nkq}^2 / \lambda_{kq})^{\rho/2}]$, where $a_{nkq} := G_n^\top \alpha_{kq}^{t+1} - (\beta_k^{t+1})^\top G \alpha_{kq}^{t+1}$. We optimize via projected gradient descent.

KGGMM for Abnormality Detection. We use the KGGMM with a small number of mixture components K, such that each component k models a significant fraction of the data, i.e., ω_k are *not* close to zero and comparable for different components k. After KGGMM fitting, we define a decision boundary $\mathcal{B}$ enclosing the normal class by a threshold τ on the minimum Mahalanobis distance across all K mixture components, such that, for a chosen component k, 98.5% of the probability mass lies within $\mathcal{B}$. τ varies with ρ; for the univariate Gaussian ($\rho = 2$) and variance σ^2, τ limits the distance to 2.5σ from the component-k mean. For the univariate generalized Gaussian, τ can be computed via the inverse cumulative distribution function that is known analytically. Because τ relies on Mahalanobis distance that is independent of scale, τ naturally extends to the multivariate case. Thus, $\mathcal{B}$ is set automatically via ρ and θ.

3 Results and Discussion

We evaluate our method for abnormality detection on simulated data and 4 large publicly available medical image datasets. Indeed, the training, i.e., model learning, for abnormality detection methods relies solely on data from the normal class, which includes outliers and mislabeled data incorrectly labeled to the normal class. We compare our KGGMM method with 7 other methods: (i) KGG, which is a special case of KGGMM with $K = 1$, (ii) standard KPCA [14], which is a special case of KGG when $\rho = 2$, (iii) Huang et al.'s RKPCA [5], (iv) one-class regularized kernel SVM [13], (v) regularized kernel SVDD [15], (vi) 2-class regularized kernel SVM, and (vii) CHLOE: a software tuned for outlier detection in images [9]. We use cross validation to tune free parameters underlying all methods, i.e., concerning the kernel, ρ (for KGGMM), and SVM regularization.

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 (Fig. 1) from a Gaussian mixture having $K = 2$ components (normal class): mean $(0, 5)$ and $(5, 0)$, modes of variation as the cardinal axes, and standard deviations along the modes of variation as $(0.25, 1.4)$ and $(1.4, 0.25)$. We then contaminate the data with outliers of 2 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. The kernel is the Euclidean inner-product. Our KGGMM learning (with $K = 2$, $\rho = 0.6$) is far more robust to outliers, with a classification accuracy of 93%, outperforming (i) KGG ($K = 1$, $\rho = 0.6$; accuracy 77%), (ii) KPCA (accuracy 54%), (iii) SVDD (accuracy 70%), and (iv) 2-class SVM (accuracy 38%).

Results on Real-World Medical Image Data. We use 4 large publicly available image datasets. We use 2 retinopathy datasets: Messidor (www.adcis.net/en/Download-Third-Party/Messidor.html; Fig. 2) and Kaggle (www.kaggle.com/c/diabetic-retinopathy-detection; Fig. 3). We use 2 endoscopy datasets for cancer detection: chromoendoscopy in Gastric cancer (aidasub-chromogastro.grand-challenge.org; Fig. 5) and confocal laser endomicroscopy in Barrett's esophagus (aidasub-clebarrett.grand-challenge.org; Fig. 6) comprising normal images including intestinal metaplasia and 2 kinds of abnormal images including dysplasia (potentially leading to cancer) and neoplastic mucosa (advanced stage cancer). The figures show that all 4 datasets, even though carefully constructed, already have outliers in the normal class. We use

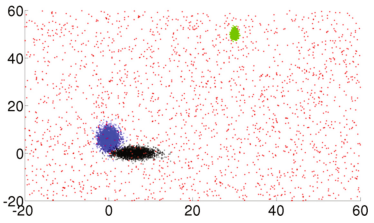


Fig. 1. Results on simulated data. Data from a 2D Gaussian mixture model (2 components, shown in blue and black) contaminated with outliers (red and green).

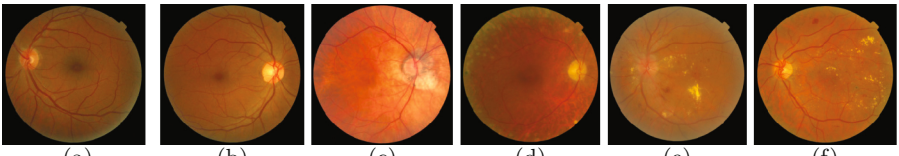
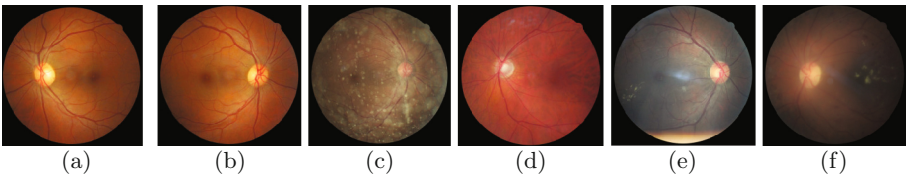
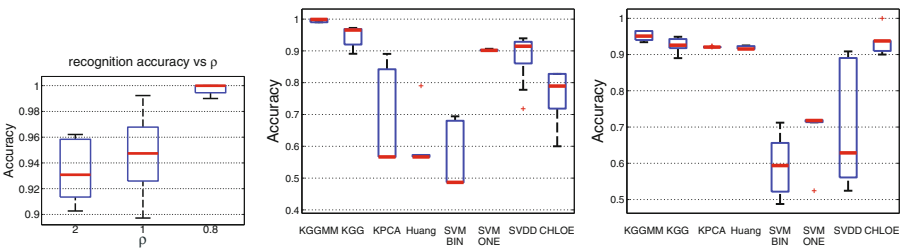


Fig. 2. Retinopathy data: Messidor. (a)–(b) Normal images. **(c)–(d)** Images labeled normal, but are outliers. **(e)–(f)** Abnormal images.



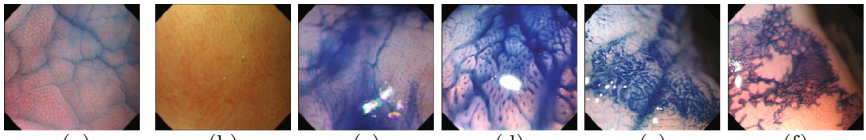
(a) (b) (c) (d) (e) (f)

Fig. 3. Retinopathy data: Kaggle. (a)–(b) Normal images. **(c)–(d)** Images labeled normal, but are outliers. **(e)–(f)** Abnormal images.



(a) KGGMM only. (b) Messidor Data. (c) Kaggle Data.

Fig. 4. Results on retinopathy data. Classification accuracy after learning on training sets contaminated with outliers, for: **(a)** KGGMM, Messidor, varying ρ, ($\rho = 2$ is Gaussian); **(b)** all methods, Messidor; **(c)** all methods, Kaggle. The box plots show variability in accuracy with resampling (uniform random) training data (20 repeats).

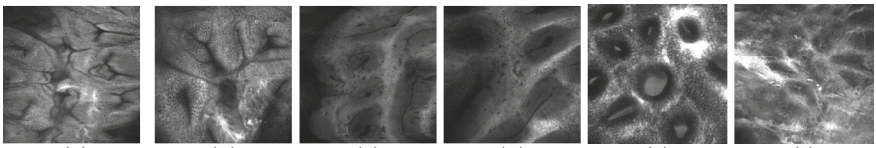


(a) (b) (c) (d) (e) (f)

Fig. 5. Chromoendoscopy data: gastric cancer. (a)–(b) Normal images. **(c)–(d)** Images labeled normal, but are outliers. **(e)–(f)** Abnormal images.

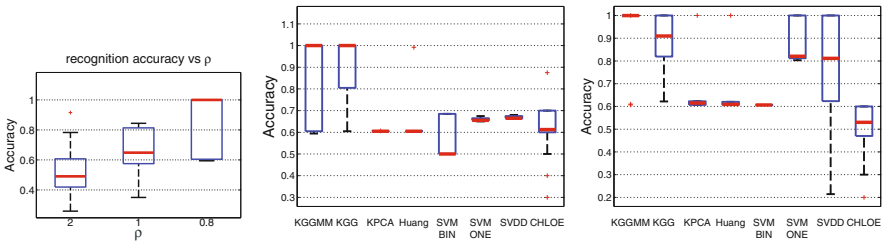
the texton-based histogram feature, using patches (9×9) to compute textons, to classify regions (50×50) as normal or abnormal. We use the intersection kernel [1]. From each dataset, we select training sets with 12000 normal image regions and, to mimic a clinical scenario, contaminate it by adding another 5–10% of abnormal image regions mislabeled as normal. The test set has 8000 normal and 5000 abnormal images. KGGMM performs best when $\rho < 1$ in retinopathy (Fig. 4(a)) and endoscopy datasets (Fig. 7(a)). The abnormality-detection accuracy of KGGMM is significantly more than all other methods for retinopathy (Fig. 4(b)–(c)) and endoscopy data (Fig. 7(b)–(c)). In almost all cases, KGGMM (we use $K = 2$ for model simplicity) performs better than KGG.

Conclusion. We have proposed a novel method for *robust kernel-based* statistical learning that relies on the generalization of the *multivariate generalized Gaussian*



(a) (b) (c) (d) (e) (f)

Fig. 6. Confocal endoscopy data: barrett's esophageal cancer. (a)–(b) Normal images. **(c)–(d)** Images labeled normal, but are outliers. **(e)–(f)** Abnormal images.



(a) KGGMM only. (b) Gastric Cancer. (c) Esophageal Cancer.

Fig. 7. Results on endoscopy data. Classification accuracy after learning on training sets contaminated with outliers, for: **(a)** KGGMM, gastric cancer, varying ρ, ($\rho = 2$ is Gaussian); **(b)** all methods, gastric cancer; **(c)** all methods, esophageal cancer. Box plots show accuracies with resampling (uniform random) training data (20 repeats).

to RKHS for *mixture modeling*. We fit our KGGMM using EM, using solely the Gram matrix. We exploit KGGMM, including *covariance* operators, for abnormality detection in medical applications where a (small) fraction of training data is inevitably contaminated because of outliers and mislabeling. The results on 4 large datasets, in retinopathy and cancer, shows that KGGMM outperforms one-class classification methods (KPCA, one-class kernel SVM, kernel SVDD), 2-class kernel SVM, and software tuned for outlier detection in images [9].

References

1. Barla, A., Odone, F., Verri, A.: Histogram intersection kernel for image classification. In: IEEE International Conference on Image Processing, vol. 3, pp. 513–516 (2003)
2. Debruyne, M., Verdonck, T.: Robust kernel principal component analysis and classification. Adv. Data Anal. Classif. **4**(2), 15167 (2010)
3. Fritsch, V., Varoquaux, G., Thyreau, B., Poline, J.B., Thirion, B.: Detecting outliers in high-dimensional neuroimaging datasets with robust covariance estimators. Med. Imag. Anal. **16**(7), 1359–1370 (2012)
4. Hoffmann, H.: Kernel PCA for novelty detection. Pattern Recog. **40**(3), 863 (2007)
5. Huang, H., Yeh, Y.: An iterative algorithm for robust kernel principal component analysis. Neurocomputing **74**(18), 3921–3930 (2011)
6. Kwak, N.: Principal component analysis by Lp-norm maximization. IEEE Trans. Cybern. **44**(5), 594–609 (2014)

7. Li, Y.: On incremental and robust subspace learning. Pattern Recog. **37**(7), 1509–1518 (2004)
8. Lu, C., Zhang, T., Zhang, R., Zhang, C.: Adaptive robust kernel PCA algorithm. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 6, pp. 621–624 (2003)
9. Manning, S., Shamir, L.: CHLOE: a software tool for automatic novelty detection in microscopy image datasets. J. Open Res. Soft. **2**(1), 1–10 (2014)
10. Mourao-Miranda, J., Hardoon, D., Hahn, T., Williams, S., Shawe-Taylor, J., Brammer, M.: Patient classification as an outlier detection problem: an application of the one-class support vector machine. Neuroimage **58**(3), 793–804 (2011)
11. Norousi, R., Wickles, S., Leidig, C., Becker, T., Schmid, V., Beckmann, R., Tresch, A.: Automatic post-picking using MAPPOS improves particle image detection from cryo-EM micrographs. J. Struct. Biol. **182**(2), 59–66 (2013)
12. Novey, M., Adali, T., Roy, A.: A complex generalized Gaussian distribution-characterization, generation, and estimation. IEEE Trans. Sig. Proc. **58**(3), 1427–1433 (2010)
13. Scholkopf, B., Platt, J., Shawe-Taylor, J., Smola, A., Williamson, R.: Estimating the support of a high-dimensional distribution. Neural Comp. **13**(7), 1443 (2001)
14. Scholkopf, B., Smola, A., Muller, K.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Comp. **10**(5), 1299–1319 (1998)
15. Tax, D., Duin, R.: Support vector data description. Mach. Learn. **54**(1), 45–66 (2004)
16. Varma, M., Zisserman, A.: A statistical approach to material classification using image patch exemplars. IEEE Trans. Pattern Anal. Mach. Intell. **31**(11), 2032–2047 (2009)