Foreign object detection and removal to improve automated analysis of chest radiographs

Laurens Hogeweg,[a)] Clara I. Sánchez, Jaime Melendez, and Pragnya Maduskar
Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre, Nijmegen 6525 GA, The Netherlands

Alistair Story and Andrew Hayward
University College London, Centre for Infectious Disease Epidemiology, London NW3 2PF, United Kingdom

Bram van Ginneken
Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre, Nijmegen 6525 GA, The Netherlands

(Received 15 August 2012; revised 29 April 2013; accepted for publication 30 April 2013; published 3 June 2013)

Purpose: Chest radiographs commonly contain projections of foreign objects, such as buttons, brassier clips, jewellery, or pacemakers and wires. The presence of these structures can substantially affect the output of computer analysis of these images. An automated method is presented to detect, segment, and remove foreign objects from chest radiographs.
Methods: Detection is performed using supervised pixel classification with a kNN classifier, resulting in a probability estimate per pixel to belong to a projected foreign object. Segmentation is performed by grouping and post-processing pixels with a probability above a certain threshold. Next, the objects are replaced by texture inpainting.
Results: The method is evaluated in experiments on 257 chest radiographs. The detection at pixel level is evaluated with receiver operating characteristic analysis on pixels within the unobscured lung fields and an A_z value of 0.949 is achieved. Free response operator characteristic analysis is performed at the object level, and 95.6% of objects are detected with on average 0.25 false positive detections per image. To investigate the effect of removing the detected objects through inpainting, a texture analysis system for tuberculosis detection is applied to images with and without pathology and with and without foreign object removal. Unprocessed, the texture analysis abnormality score of normal images with foreign objects is comparable to those with pathology. After removing foreign objects, the texture score of normal images with and without foreign objects is similar, while abnormal images, whether they contain foreign objects or not, achieve on average higher scores.
Conclusions: The authors conclude that removal of foreign objects from chest radiographs is feasible and beneficial for automated image analysis.
[http://dx.doi.org/10.1118/1.4805104]

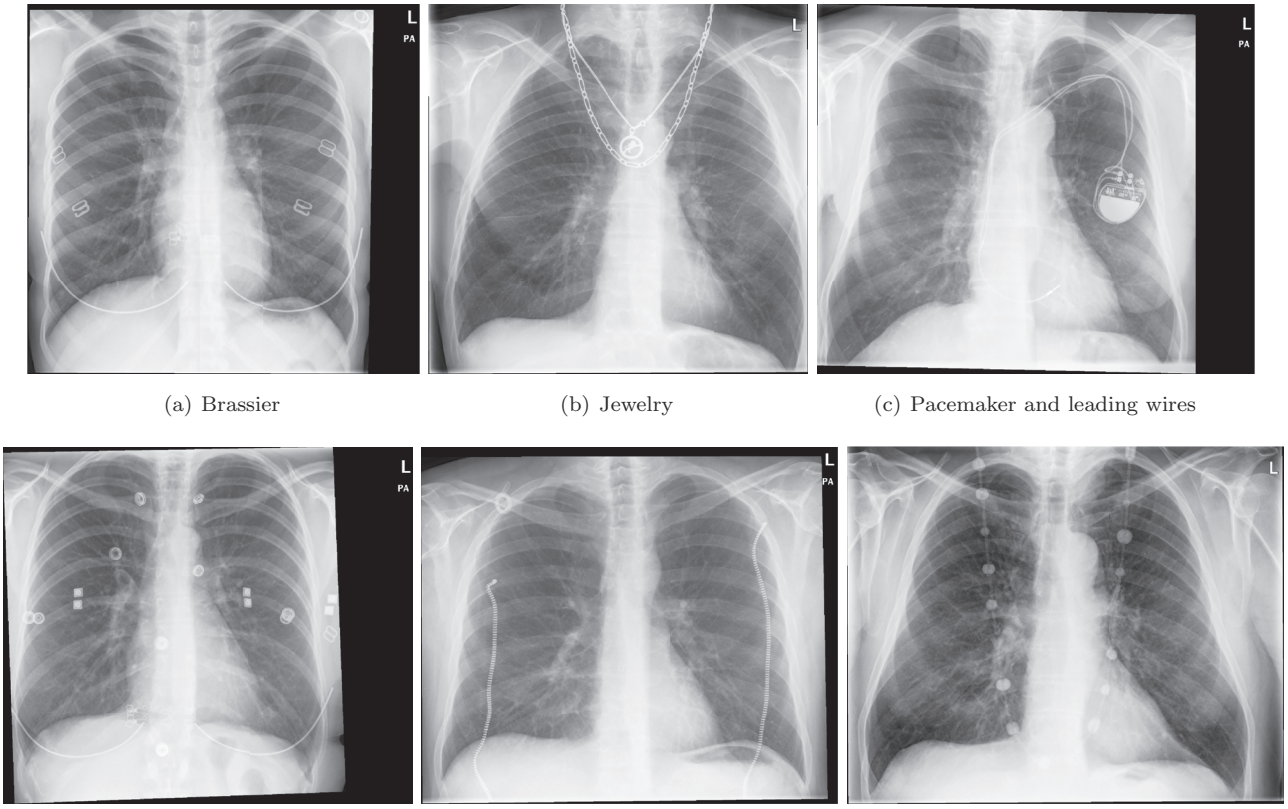
Key words: foreign objects, pixel classification, detection, chest radiography, artifact restoration, CAD

I. INTRODUCTION

Algorithms for automated analysis of images often require a minimum quality of the input. Scientific studies that evaluate the performance of automated image analysis algorithms often exclude images of poor quality, either by removing them manually from the dataset or by using an automated algorithm to determine if the image quality is sufficient. Such quality assessment algorithms have been proposed for, e.g., retinal images[1] and fingerprints.[2]

In some situations detection and exclusion of low quality images may be sufficient, for example when it is possible to directly request a new acquisition. However, it will often be unfeasible or unacceptable to acquire a new image. An example is a screening program where participants might not be motivated to have a second exam taken. Bedside radiographs often contain foreign objects, such as catheters, which cannot be removed before acquisition. It may also be the case that image quality is sufficient for reading by human experts, who can readily ignore and dismiss the artifacts, but automated analysis by computers will produce false alarms. If the locations of artifacts are known, one could attempt to deal with the issue by ignoring the computer output in areas affected by the artifact. A more ambitious strategy is to try to remove the artifacts. In this work, we attempt to remove the artifacts in order to approximate the unaffected image as closely as possible.

Our focus is on detection and removal of foreign objects in chest radiographs. We use data from a large-scale screening program for tuberculosis. In a representative sample of 1000 chest radiographs taken from this screening program almost 20% of the lung fields contained one or more foreign objects. Figure 1 shows a number of different types of objects that were found to be present in these images. People are asked to remove their coats and any heavy chains, but are not required to fully disrobe; disrobing is time consuming and for some participants a barrier to participate.



(a) Brassier (b) Jewelry (c) Pacemaker and leading wires

(d) Buttons and brassier (e) Zippers (f) Jewelry

FIG. 1. Foreign objects in chest radiographs from Find and Treat database.

To the best of our knowledge the topic of foreign object removal in chest radiographs has not been investigated before. The most closely related work in chest radiography image analysis has been done on automatic detection of catheter tips[3] (without the aim of actually removing the catheters) and the removal of anatomical structures such as ribs and clavicles by suppressing them.[4,5]

More generally, the restoration of digital images has received considerable attention in the recent literature. Many older studies deal with reverting the effects of noise such as blurring and speckle.[6,7] For the purpose of this paper, we are especially interested in methods that are able to fill large holes (left by the foreign objects) in the image. Lee *et al.*[8] used bilinear interpolation along the short axis of elongated holes to remove hairs from dermoscopy images. One of the first automatic methods that uses texture synthesis to fill holes is the work of Efros and Leung[9] who employed nonparametric sampling. Bertalmio *et al.*[10] improved on this method by employing techniques derived from inpainting. Inpainting is a concept that originates from the restoration of paintings where the goal is to repair damaged parts in a visually convincing way.[11] Digital inpainting is described as the process where in a first step structural elements are continued into the holes, subsequently color is added to the still missing areas and finally texture is added. Bertalmio *et al.*[10] used anisotropic diffusion to propagate linear structures along isophotes. In an improved version texture was also added.[12] Criminisi *et al.*[13] further improved on this approach by removing the need to separate the image into its structural and texture component before inpainting. Both methods rely on nonparametric texture sampling to synthesize missing texture. Texture sampling can also be used as a stand alone inpainting technique if the restoration of structural elements is of less importance.

The work most similar to our paper, in its aim to improve automatic processing of medical images by repairing artifacts, has been done in the field of dermoscopy. In dermoscopic images linear structures such as hairs and rulers complicate (automated) analysis. Zhou *et al.*[14] used automatic line extraction followed by inpainting to remove the objects. Wighton *et al.*[15] used a similar approach but focused on the comparison between two methods to remove simulated hair. They found that using a specific type of inpainting, so called exemplar based inpainting, yielded images more similar to the original unaffected ones than using the linear interpolation algorithm DullRazor.[8] In a more recent comparison it was found that fast marching inpainting[16] performed better than linear interpolation, nonlinear diffusion and exemplar-based inpainting in hair removal judged by segmentation performance and texture analysis.[17] In this work, we employ a modified version of the texture synthesis method described by Efros and Leung[9] to inpaint areas where artifacts are superimposed on relevant image structure in chest radiographs.

The paper is organized as follows. In Sec. II, the data used in this study is described. Section III details the steps to automatically detect, segment, and remove foreign objects from chest radiographs. The experiments to evaluate the method,

TABLE I. Sets of images used for training and evaluating the algorithm.

Set	Purpose	Content
A	Training set for foreign object detection	100 normal images with 331 foreign objects
B	Evaluation set for foreign object detection	107 normal images with 271 foreign objects
C	Training set for texture analysis	90 normal images, 48 with tuberculosis related pathology
D	Evaluation set for texture analysis	Set B + 100 normal images without foreign objects + 50 with tuberculosis related pathology (8 containing foreign objects)

in terms of detection and segmentation performance and in terms of its effect on subsequent use of the repaired images in a CAD system, are presented in Sec. IV and result are given in Sec. V. Section VI discusses the results and Sec. VII concludes.

II. DATA

For evaluation of the detection and removal of foreign objects, a large chest radiograph database from a tuberculosis screening program for high risk groups in London was used.[18] All images were digital and acquired with a single unit (DigitalDiagnost Trixel, Philips Healthcare, The Netherlands). In this work, all images were scaled to a width of 1024 pixels, corresponding to a resolution ranging from 0.22 to 0.38 mm per pixel depending on the size of the original image. Although the original resolution is about a factor 2 higher (0.144 mm pixel size), we have found that this resolution is sufficient for the detection of tuberculosis related abnormalities in chest radiographs. This resolution is also sufficient for the detection and segmentation of foreign objects. If desired, the inpainting could also be applied to full resolution images.

Approximately 20% of the images in this database contain foreign objects. A subset of this database was used to evaluate the automatic detection of foreign objects and its effect on automatic detection of textural abnormalities. Table I lists four datasets defined for this study. Sets A and B were used to train and evaluate the detection of foreign objects (Sec. V.B), sets C and D were used to train and evaluate the detection of textural abnormalities with and without foreign object removal (Sec. V.C). Images in datasets A–D were randomly selected from the full database.

For training of the detection system, precise annotations of foreign objects in the chest radiographs are required. These were obtained semiautomatically (see Fig. 2). First, the contrast between the high density object with the background was further enhanced by local normalization[19] at a scale of $\sigma = 8$ pixels. This scale was determined by visual inspection and isolates high density objects from the background [Fig. 2(b)]. The human operator set an appropriate threshold on this image and outlined the region of a foreign object. This region was masked and connected component analysis was used to quickly select all parts of the object to be segmented. The resulting binary mask was adjusted with a paint tool to obtain a precise segmentation and, if necessary, hole filling was applied. Objects with an area smaller than 120 pixels were excluded. This prevents detection of very small objects that are usually spurious. Using this procedure, 331 and 271 foreign objects were annotated in sets A and B, respectively, ranging in area from 120 to 6316 pixels.

III. METHODS

The method starts with automatic segmentation of the unobscured lung fields. The proposed method consists of three stages (only performed inside the lung fields): detection of pixels belonging to foreign objects, segmentation of the objects, and removal of the object using inpainting. Both foreign

(a) input image (b) local normalization (c) segmentation (d) hole filling

FIG. 2. Semiautomatic segmentation of foreign objects in chest radiographs. (a) A part of a chest radiograph containing a foreign object. (b) Local normalization is applied to increase the contrast between the dense object and its direct surroundings. (c) The object is roughly outlined and the outlined region is thresholded and the correct connected components are selected. This segmentation is then manually postprocessed with a painting tool. (d) When deemed necessary, interior pixels are added by hole filling.

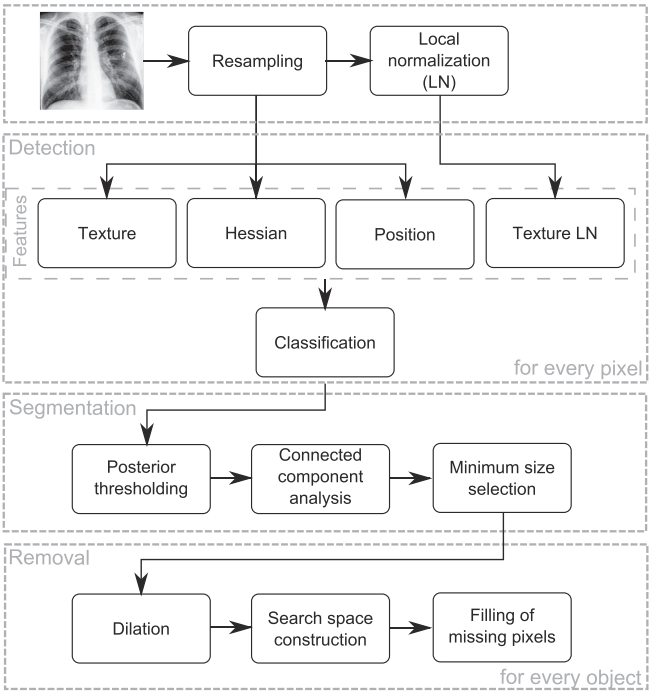


FIG. 3. Overview of method. The method starts with resampling to a width of 1024 pixels and local normalization. Foreign objects are detected by computing features for every pixel inside the lung fields, followed by supervised classification. Objects are segmented by thresholding and connected component analysis. Removal of foreign objects is performed using a texture synthesis method which fills missing pixels with values obtained from the best matching patch in a search space around the object. See the text for details of each step.

object detection and lung field segmentation use pixel classification. An overview of the method is given in Fig. 3. The following paragraphs describe the method in detail.

III.A. Lung field segmentation

The detection and removal of foreign objects is limited to the unobscured lung fields as this is the main area of interest for many other applications of automated analysis. An automatic lung segmentation algorithm based on pixel classification was used to find the lung fields.[20] The system was trained with 500 training images where lung contours were manually outlined. These images were consecutively selected and some of them contained foreign objects and/or pathology. These images were not further used in the rest of the study. A number of small changes were made with respect to the system described by van Ginneken *et al.*[20] Features were calculated at images of 1024 pixels wide (instead of 256 pixels), and resulting images are 512 pixels wide (instead of 256 pixels). As feature set, we used Gaussian derivative and Hessian based features on the original images as described in Sec. III.B.2. Pixels were classified using a k-nearest neighbor (kNN) classifier ($k = 15$). The resulting probability map was slightly blurred ($\sigma = 0.7$ pixels), thresholded at a probability of 0.5, and morphologically closed with a spherical kernel ($r = 10$ pixels). To segment the lungs, the two largest components were selected.

III.B. Detection of foreign objects

III.B.1. Pixel classification (PC)

In this methodology, the segmentation problem is recast into a pattern classification task.[21,22] A number of continuous characteristics (features) are calculated for a number of samples (positions, pixels) in an image. A classifier is trained using labeled samples from a database of training images. Examples of labels are inside/outside foreign objects and inside/outside the unobscured lung fields. The classifier provides the mapping from features to class labels. In this work, we use classifiers that provide a posterior probability that indicates how likely a sample should receive a label. Test images can be segmented afterwards by computing the features for each position and applying the classifier.

III.B.2. Features

Three types of features were calculated for each sample: texture features based on Gaussian derivatives (on original and locally normalized images), features derived from the Hessian matrix and position features. First each image is resized to a width of 1024 pixels. To capture local image structure,[23] the output of Gaussian derivative filtered images of order 0, 1, 2 (L, L_x, L_y, L_{xx}, L_{xy}, L_{yy}), at scales 1, 2, 4, 8, 16, 32, and 64 pixels were calculated. The small scales provide information about the fine image structure and the larger scales about the neighborhood of the pixel. For speed improvement, the recursive implementation described by Deriche[24] was used. Foreign objects typically have a high density compared to their background. As explained in Sec. II applying local normalization to the image will improve the contrast of the object with the background. To reflect this in the feature set, images were locally normalized with $\sigma_{LN} = 8$ pixels, the same scale as used for annotation, and features derived from the output of Gaussian derivative filtered images of order 0 and 1 (L, L_x, L_y) at scales 1 and 2 were added. The use of small scales specifically enhances strong edges, which are characteristic of foreign objects encountered in our dataset. Certain types of foreign objects consist of thin, elongated lines. Hessian matrix derived features were used to detect the presence of these line like structures.[25] Considering the two eigenvalues of the Hessian matrix λ_1, λ_2 with $|\lambda_1| > |\lambda_2|$, two measures were derived: (1) $\sqrt{(\lambda_1^2 - \lambda_2^2)}$ to extract the lineness of the local image structure and the largest absolute eigenvalue $|\lambda_1|$ to indicate the strength of the response. These Hessian features were calculated at scales 1, 2, 4, 8, and 16 pixels. Finally the x and y position were added to account for the difference in background appearance across the lung. In total 61 features were computed per sample.

III.B.3. Classification

The training set was constructed by sampling inside the lung fields 100% of the available positive (foreign object) pixels and a random 0.5% of the available negative pixels per image. Only a small percentage of normal pixels was

1. Given datapoint and label pairs $(\mathbf{x}_1, y_1), ..., (\mathbf{x}_N, y_N)$ where $\mathbf{x}_i \in X, y_i \in Y = \{-1, +1\}$
2. Start with $H(\mathbf{x}_i) = 0$ and weights $w_i = 1/N$, $i = 1, ..., N$
3. Repeat for $m = 1, ..., M$
   (a) Find the optimal weak classfier h_m for (X, Y) and current weights w_i
   (b) Update strong classifier $H(\mathbf{x}) \leftarrow H(\mathbf{x}) + h_m(\mathbf{x})$
   (c) Update weights for examples $w_i \leftarrow w_i e^{-y_i h_m(\mathbf{x}_i)}$ for $i = 1, ..., N$

FIG. 4. GentleBoost algorithm.

sampled because they occur much more frequently in the original image. To reduce computation time test images were subsampled by a factor 2 resulting in probability maps with a width of 512 pixels. For segmentation, removal, and evaluation the probability maps were resized back to a width of 1024 pixels.

All features were normalized to zero mean and unit standard deviation before classification. A number of different classifiers for foreign object detection were tested. In the following description of the classifiers, we use boldface to denote vectors or matrices, $\mathbf{x}$ for feature vectors (samples), and $y \in \{-1, +1\}$ for sample labels of classes 1 and 2. All classifiers output posterior probabilities.

Linear discriminant analysis. Fisher's linear discriminant analysis (LDA) is a discriminative method which is commonly used in CAD applications for its speed and simplicity. The method finds a linear discriminant function in a supervised fashion assuming that the class density $f_c(\mathbf{x})$ for each class c is a multivariate Gaussian distribution with the same covariance matrix Σ:

$$f_c(\mathbf{x}) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{m}_c)^T \Sigma^{-1}(\mathbf{x}-\mathbf{m}_c)\right) \quad (1)$$

with $c = 1, 2$ and $\mathbf{m}_c$ the class mean. The posterior probability for $y = 1$ given $\mathbf{x}$ is given by

$$P(y=1|\mathbf{x}) = \frac{\pi_1 f_1(\mathbf{x})}{\pi_1 f_1(\mathbf{x}) + \pi_2 f_2(\mathbf{x})}, \quad (2)$$

where π_c is the prior probability of class c.

Nearest mean classification. Nearest mean classification assigns to a test point the label of the nearest class mean. The posterior probability is calculated in a similar way as LDA, using Eq. (2), under the assumptions of multivariate Gaussian class distributions with the same covariance matrix Σ and $\Sigma = \sigma^2 I$ where I is the identity matrix. Like LDA, the nearest mean method is unable to model complex decision boundaries but it is fast, performs well in some settings, and makes no assumptions on the data.

k-nearest neighbor. The kNN classification is a nonparametric method where the decision boundary is constructed locally in an area around the query sample. The posterior probability for $y = 1$, given $\mathbf{x}$ is given by

$$P(y=1|x) = \frac{k_1}{k}, \quad (3)$$

where k_1 is the number of samples among the k nearest neighbors with label $y = 1$. The Euclidean distance was used as a distance measure in this work. kNN has the attractive property that with increasing training size, the conditional error approaches the Bayes error.[22] To speed up the classification, the tree-based implementation by Arya *et al.*[26] was used. This implementation uses an approximate solution controlled by the variable ϵ, which ensures that the approximate nearest neighbors are no more than $(1+\epsilon)$ times the distance away from the query point than the actual nearest neighbors. ϵ was set to 2 in this work.

GentleBoost. The GentleBoost algorithm belongs to the family of ensemble classifiers and was described by Friedman *et al.*[27] A number of weak classifiers h_m are sequentially combined into a strong classifier H using the algorithm shown in Fig. 4. The algorithm uses adaptive Newton steps in each round m to minimize the weighted squared error

$$J_m = \sum_{i=1}^{N} w_i (y_i - h_m(\mathbf{x}_i))^2, \quad (4)$$

where $w_i = e^{-y_i H(\mathbf{x}_i)}$ are the weights, h_m the weak classifier, and N the number of training samples. The optimal weak classifier is then added to the strong classifier H and the weights are adapted. For the weak classifier, any suitable algorithm can be chosen, but for GentleBoost often a simple algorithm, such as regression stumps, is used. Regressions stumps are basically decision trees with one node and are defined as $h_m(\mathbf{x}_i) = a\delta[\mathbf{x}_i^f > \theta] + b\delta[\mathbf{x}_i^f < \theta]$, where f is the the feature number and δ the indicator function. The stump is optimized by finding the parameters $\{a, b, f, \theta\}$ that minimize J_m. A closed form solution for a and b can be derived and $\{f, \theta\}$ are found by exhaustive search.[28] The trained strong classifier H gives the log-odds of being in class y, where $H(\mathbf{x}) = \log P(y|\mathbf{x})/P(-y|\mathbf{x})$, $y \in \{-1, +1\}$. The posterior probability for $y = 1$ given $\mathbf{x}$ is estimated using a sigmoid function as follows:

$$P(y=1|\mathbf{x}) \approx \frac{1}{1+e^{-H(\mathbf{x})}}. \quad (5)$$

GentleBoost has been shown to have improved performance compared to other classifiers such as AdaBoost.[29] For boosting algorithms in general, similar performance was found as for neural networks but with decreased training times.[30]

Support vector machine. The support vector machine (SVM) constructs a hyperplane in a high-dimensional space in such a way that the distance between the hyperplane and the two classes is maximal, which makes it a maximum-margin classifier. The hyperplane is found by solving the following minimization problem[31]

$$\begin{aligned} \underset{\mathbf{w},b,\xi}{\text{minimize}} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & y_i(\mathbf{w}^T\phi(\mathbf{x}_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, \end{aligned} \tag{6}$$

where $\mathbf{w}$ and b are the weights and the bias of the hyperplane, respectively. The parameter $C > 0$ controls the misclassification error induced by the slack variables ξ, which are introduced to allow for solutions when a hyperplane splitting the classes does not exist. The function ϕ maps the feature vectors into a higher dimensional space where a hyperplane may be easier to find. ϕ is related to a kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T\phi(\mathbf{x}_j)$. Different kernel functions can be used. Setting $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T\mathbf{x}_j$ gives a linear kernel. The Gaussian radial basis function $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma||\mathbf{x}_i - \mathbf{x}_j||^2)$, $\gamma > 0$ is often used as a nonlinear kernel and introduces an extra parameter γ. SVM does not provide posterior probabilities $P(y = 1|\mathbf{x})$ directly, but these can be estimated from the distances of samples to the hyperplane by cross validation on the training set, for details see Chang and Lin.[32] Training times for SVM can be considerable as C and γ have to be determined in an exhaustive search procedure. Testing times are typically much faster and related to the number of support vectors needed to define the hyperplane.

III.C. Segmentation

The output of the pixel classifier is a probability for each location in the lung fields to belong to a foreign object. A binary segmentation is obtained by thresholding the probabilities of the pixel classifier using a threshold p_t. Subsequently connected component analysis (using eight-connectedness) is performed. Only objects with an area >120 pixels were retained. The effect of p_t on the detection performance is evaluated in Sec. V.B.

III.D. Removal of foreign objects

After detection, foreign objects were removed from the image to restore the appearance of the background lung. For removal, we adopted a texture synthesis algorithm, which uses non-parametric sampling to recover the texture at the location of the removed foreign object. The removal is performed on images of 1024 pixels wide. Incorrectly segmented foreign object pixels at the boundary (false negatives), which typically have a high density, could be incorrectly used as example pixels for the removal algorithm and lead to artifacts in the restored image. To reduce this risk the detected objects are slightly dilated with 4 pixels before texture synthesis is performed.

For texture synthesis the method of Efros and Leung[9] was taken as the basis. Missing pixels are filled one at a time by finding pixels with similar neighborhoods and copying the value of the pixel with the best matching neighborhood. The neighborhood is defined as a square patch surrounding the pixel. We used patches of 11×11 pixels. Similar neighborhoods are determined by calculating distances between the patch for the missing pixel and all other possible neighborhoods. The distance d is defined as the sum of squared differences (SSD) between the two patches. The SSD is only calculated using pixels that are known in both patches. To prevent the method from getting stuck in local minima, a random patch is selected from a set of best matching patches with $d < (1 + e)d_{best}$, where d_{best} is the distance of the best matching patch and e is a threshold.

While the method is intuitively simple, it is in practice very slow because the whole image has to be searched for every missing pixel. Three modifications were made to make the method useful in practice. To speed up the search process of similar patches we used approximate nearest neighbor (ANN) search,[26] a fast version of kNN search (inspired by the use of tree structured vector quantization in Wei and Levoy[33]). The use of kNN search also replaces the parameter e from the original algorithm, which gives a variable number of best matching patches, by a fixed number of patches controlled by k. We used $k = 10$. In a structured image such as a chest radiograph, similar patches are expected to be found closer to the missing pixel, therefore we limited the search area to an area of 50 pixels around each of the objects to be filled. The considerable reduction of the size of the search space also improves computation times. Contrary to the original algorithm, which updates the search space after each iteration, the search space is only constructed once per object. Finally we took the output of the DullRazor[8] method as a pre-processed input image for texture synthesis. DullRazor performs bilinear interpolation at each missing pixel between the endpoints of the shortest ray crossing the hole. Eight rays were cast in equally spaced directions to determine the shortest ray. Prefilling the hole is necessary as kNN requires complete patches (with no missing pixels) to search with, while missing pixels typically also have missing pixels in their neighborhood.

IV. EXPERIMENTS

IV.A. Foreign object detection: Pixel based evaluation

The training set consisted of 59 887 pixels (49 268 pixels from background, 10 619 pixels from foreign objects) sampled from set A. The test set consisted of 108 052 pixels (47 743 pixels from background, 60 309 pixels from foreign objects) sampled from set B.

For detection of foreign objects, we tested LDA, kNN classification, Nearest mean classification, SVM, and GentleBoost. Optimal parameters and features were determined for the classifiers. From pilot experiments, it was determined that a high pixel level specificity is required to limit the number of false positive object detections. Therefore the optimization criterion was set to the sensitivity at 0.995 specificity. The

value of k in kNN was optimized in cross validation on the training set. Odd values of k in the range 1–301 were tested, and $k = 19$ was determined to be the optimal value. For SVM, a Gaussian kernel function was used, the hyperparameters C and γ were optimized in a grid search by cross validation on the training set, according to the recommendations in Hsu *et al.*[34] The optimal values of C and γ were 2^{-1} and 2^{-5}, respectively. GentleBoost used regression stumps as the weak classifier. A number of 1000 stumps were added to train the classifier. Feature selection was performed using sequential forward selection (SFS) (Ref. 35) for each of the classifiers. Feature selection for SVM was not performed, because optimal values for the hyperparameters would have to be determined for every different subset of features, leading to prohibitively long computation times.

Receiver operating characteristic (ROC) analysis was performed on the test set. The pixel based classifier performance was measured using the area under the ROC curve A_z. Differences between A_z values were determined with bootstrapping,[36] using 1000 bootstrap samples and a significance level $\alpha = 0.05$.

IV.B. Foreign object detection: Object based evaluation

Object detection performance is determined using free response operator characteristic (FROC) analysis.[37] For FROC construction criteria are needed for true positive (TP), false positive (FP), and false negative (FN) objects. The criteria are based on the fraction of positive pixels detected $\Omega = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$. An object in the segmentation is TP when it overlaps with a foreign object of which at least 50% of the pixels are detected ($\Omega > 0.50$). If an object overlaps but less than 50% of the pixels are detected ($0 < \Omega < 0.50$), it is considered a FN. Objects in the reference standard which do not overlap with any object in the segmentation are also FN. Finally, an object in the segmentation is considered FP when $\Omega = 0$.

During construction of the FROC curve, p_t is lowered to obtain segmentations at different levels and determine the number of TP, FN, and FP objects. The area of the segmented objects (connected components) typically grows when the threshold p_t is lowered, and may lead to merging of two or more FP objects into one and a counterintuitive reduction of number of FPs at a lower value of p_t. This has been noted as an issue in FROC analysis in several studies.[38,39] To prevent this effect, a local maxima detection scheme is used. FROC construction starts at $p_t = 1.0$. At each level of p_t, FP objects are identified in the segmentation. For each FP object it is determined whether it overlaps with an object detected at a higher p_t. If it was previously detected it is ignored, otherwise the object is recorded as FP and assigned a score equal to the current p_t. In a similar way the TP and FN objects are assigned a score based on the level of p_t where they first appeared. In the reported results, the step size for p_t was 0.05.

The object score is used to construct the FROC curve. Pilot results showed that many false positive responses of PC occur at the boundary of the lung where the high gradient transition between bone and lung tissue mimics that of foreign object and lung tissue. To reduce the effect of lung segmentation inaccuracies on the results, the region within a distance of 4 pixels (0.88–1.52 mm) from the automatically detected lung boundary was excluded from analysis.

IV.C. Effect on textural abnormality detection

One of the goals of foreign object removal is to improve automated analysis of chest radiographs. Textural (parenchymal) abnormalities are a common disease pattern in chest radiographs and can be found in different types of diseases, such as pneumonia or tuberculosis. To evaluate the effectiveness of the foreign object detection and removal algorithm, a previously described CAD system for tuberculosis detection on chest radiographs was used.[40,41] Classification is based on texture analysis of small circular image patches placed inside automatically detected lung fields.[20] Features based on moments from Gaussian derivative filtered images are calculated for each patch and assigned a probability of abnormality using LDA. One texture score, reflecting the total load of abnormal patches in the image, was determined by integrating the individual patch probability scores. This integration was performed by calculating the 95th percentile of the cumulative histogram of patch probabilities[42] which provides a robust estimate of the relative area of lung affected by pathology.

The textural abnormality system was trained with images in set C. Textural abnormalities in this set were outlined to provide examples of abnormal patches. Examples of normal patches were only sampled from normal images. The total number of normal and abnormal patches used to train the LDA classifier was 122 932 and 4958, respectively. The system was then applied to test images in set D.

The CAD system was applied two times; first using features derived from original test images and then from processed test images where foreign objects had been removed. Images in set D were divided into three groups: normal images containing foreign objects, abnormal images, and normal images without foreign objects. Texture scores were calculated for the six situations (original and processed images in the three groups) and displayed using box plots. Differences between means of the groups were compared using unpaired Student's t-tests ($\alpha = 0.05$).

V. RESULTS

V.A. Foreign object detection: Pixel based evaluation

All classifiers achieved a high pixel classification performance using the same full set of computed features. Differences are small, but significant differences in A_z were found between LDA, the best performing classifier, and the other classifiers. At high specificities (0.99–1.0), the operating region for segmentation, kNN and SVM are the best classifiers [Fig. 5(b)]. Nearest mean classification had the lowest performance but still reached a high overall A_z value. Feature selection did not lead to improved sensitivity at the level of 0.995 specificity, except for nearest mean classification, which

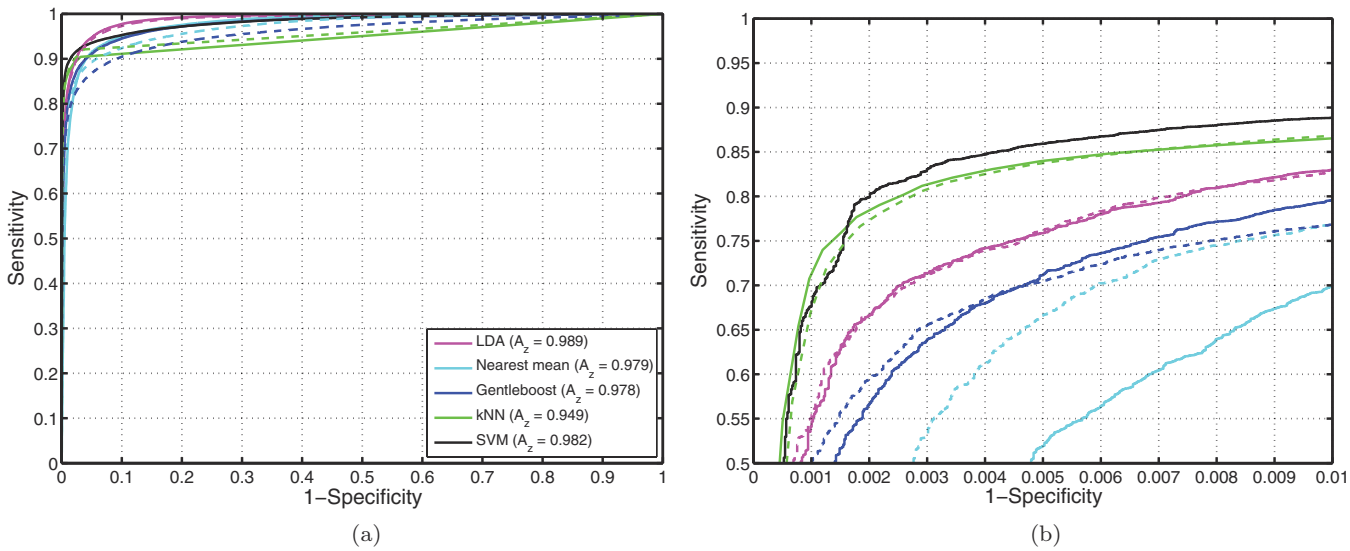


FIG. 5. ROC Analysis of pixel classification. Shown are results for linear discriminant analysis (LDA), Nearest mean, k-nearest-neighbor (kNN), GentleBoost, and SVM. The results with feature selection are indicated with dashed curves. (a) Full ROC curve. The area under the ROC curve A_z is displayed in the legend. (b) Zoomed in ROC curve at high specificities where the pixel classifier operates.

remained the worst performing classifier. Despite the lack of performance improvement, feature selection provides insight into how individual features perform. The five features performing best individually and the best subset of five features with kNN as classifier are shown in Table II. The features that were most often selected, namely λ_1 and second order derivatives at small scales, reflect the presence of fine line-like structures. This is due to the fact that foreign objects are mainly thin elongated structures with sharp borders, specially highlighted at small scales. Features at larger scales also contributed to classification but mostly in combination with other features.

V.B. Foreign object detection: Object based evaluation

Figure 6 shows FROC curves of the tested classifiers indicating object detection performance for the 0.0625-4.0 FP/image operating range. The left most point of a curve indicates the detection performance for the lowest probability map threshold ($p_t = 0.05$).

kNN showed superior sensitivities compared to the other classifiers in the analyzed FP/image range. SVM was the second best performing classifier. LDA showed worse performance than kNN and SVM. GentleBoost has reduced sensitivities compared to kNN, LDA, and SVM; while nearest mean classification has markedly reduced performance compared to all classifiers. For the remaining experiments, we used kNN. At a level of 0.25 FP/image 95.6% of the objects were successfully detected using kNN. This level corresponds to $p_t = 0.90$ and was used in subsequent experiments unless indicated otherwise. At $p_t = 0.90$, the corresponding per pixel sensitivity and specificity are 0.65 and 0.999, respectively. This operating point corresponds to the operating range where kNN performs best on the per pixel level [Fig. 5(b)]. Increasing the sensitivity at the object level of kNN beyond 95% finds only slightly more foreign objects at the expense of a large increase in the number of FP objects.

V.C. Effect on textural abnormality detection

Figures 7 through 9 show an overview of the results of the detection, removal, and effect on texture analysis of 12 selected cases. Figures 7 and 8 show cases containing foreign objects, Fig. 9 cases without foreign objects. The fourth example in Fig. 7 displays an abnormal case containing extensive

TABLE II. Analysis of feature performance on the pixel level with kNN as classifier. On the left are shown the five features which perform best individually. On the right the five features first selected using sequential forward selection (SFS), in other words the best performing subset, are shown.

No.	Individual performance	Sensitivity at 0.995 specificity	Selection order in SFS	Sensitivity at 0.995 specificity
1	λ_1 ($\sigma = 1$ pixels)	0.356	λ_1 ($\sigma = 1$ pixels)	0.356
2	L_{yy} ($\sigma = 2$ pixels)	0.316	L_{xx} ($\sigma = 4$ pixels)	0.516
3	L_{yy} ($\sigma = 4$ pixels)	0.288	L_{yy} ($\sigma = 4$ pixels)	0.643
4	λ_1 ($\sigma = 2$ pixels)	0.285	L_y ($\sigma = 64$ pixels)	0.674
5	L_{yy} ($\sigma = 1$ pixels)	0.271	L_y ($\sigma = 1$ pixels)	0.743

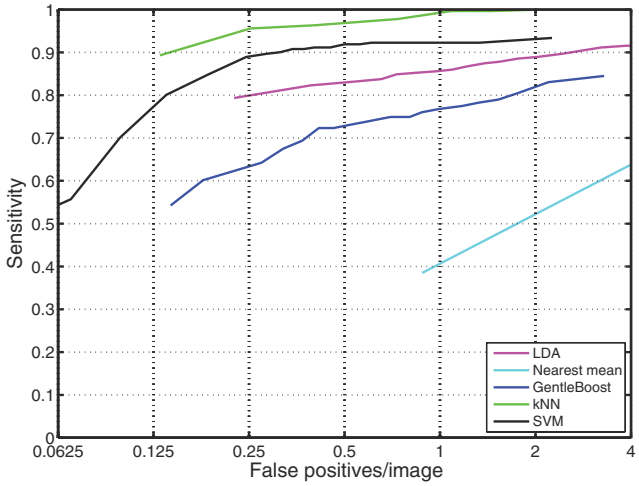


FIG. 6. Performance analysis at the object level. FROC curve for different classifiers [see Fig. 5(b)] in the 0.0625-4.0 FP/image operating range. The x axis is logarithmically scaled. The left most point of a curve indicates the detection performance for the lowest probability map threshold ($p_t = 0.05$).

pathology. From top to bottom the original images are shown, followed by the manual segmentations of the foreign objects (the reference standard). The next row shows detection using pixel classification. The fourth row shows the segmentation as produced by thresholding the output of the pixel classifier with $p_t = 0.90$, connected component analysis, and retaining components with a minimum area of 120 pixels. The fifth row shows the images with the segmented objects inpainted by texture synthesis. The last two rows show the output of the textural abnormality detection system on the original images, and on the processed images in which the foreign objects have been inpainted. Applying texture analysis to the original images leads to false positive responses due to the presence of foreign objects. The strong responses lead to texture scores in normal images that are similar to those in abnormal images. After removing the foreign object by inpainting, the false positive responses have mostly disappeared, leading to markedly reduced texture scores. In cases containing no foreign objects the texture response is identical before and after removal, except for an occasional false positive response on the lung border. Close-ups of a number of foreign objects are shown in Fig. 10.

To quantify the effect of foreign object removal, CAD texture scores were compared between normal and abnormal images and before and after image restoration in set D. Boxplots[43] in Fig. 11 show that texture scores in normal images with foreign objects are higher than in normal images without foreign objects before processing (boxplots 1 and 5; $p < 0.05$). This reduces the ability of the textural abnormality system to discriminate between abnormal images and normal images with foreign objects as they show no significant difference in scores (boxplots 1 and 3; $p = 0.35$). After processing, the scores of normal images with foreign objects are reduced to a similar levels as normal images without foreign objects (boxplots 2 and 6; $p = 0.93$). The restoration has no noticeable effect on texture scores of images containing no foreign objects (boxplots 5 and 6; $p = 0.89$). A slight, but not significant, reduction is observed in the scores of abnormal images as some of them contain foreign objects (boxplots 3 and 4; $p = 0.63$). Further analysis of the abnormal cases showed a similar pattern as in the normal cases. After processing, the average score of the eight abnormal cases containing foreign objects scores was reduced to a similar value as the average score of the 42 abnormal cases without foreign objects before processing ($p = 0.82$). This indicates that, although scores in abnormal cases are reduced, it can be predominantly ascribed to the foreign objects being removed, not to the removal of pathology. An example of an abnormal case with a zipper removed can be seen in the 4th column of Fig. 7.

VI. DISCUSSION

A method was presented to automatically detect and remove foreign objects in chest radiographs and was evaluated on images acquired from a tuberculosis screening program. In screening practice, the occurrence of foreign objects is not uncommon and automatic removal is a necessary prerequisite for other automatic processing of chest radiographs. The contribution of this paper is two-fold: (1) the application of state-of-the-art techniques for segmentation and image restoration to an unexplored application and (2) the modification of an existing texture synthesis method which reduces computation time greatly and renders its use more practicable. In this section, the detection and segmentation performance in relation to the false negatives and false positives are discussed first. Then a discussion of the removal of the objects and its usefulness in practice follows. Finally the use of the system in a practical context is discussed.

Detection and segmentation of the foreign objects was in general quite accurate. Clearly delineated objects, such as metal objects (e.g., brassieres, coins, keys) were typically detected, accurately segmented, and removed by the algorithm. This is illustrated by the first three close-ups of foreign objects in Fig. 10 where after inpainting the previously affected area is largely artifact free. The last close-up of Fig. 10 indicates that a perfect segmentation is not required to obtain a convincing inpainting result. A slight over-segmentation of objects is not a problem as the FP pixels will be restored by the inpainting algorithm. To make the algorithm less sensitive to under-segmentation obtained segmentations were slightly dilated before inpainting. A larger dilation might further reduce the risk of FN pixels, but it will also make it difficult for the inpainting to succeed as larger holes lead to information loss about the local appearance.

The algorithm missed 12/271 (4.4%) of the foreign objects at the cutoff used for segmentation. The majority of them were small (10 FN objects with area <400 pixels, 20–50 mm^2 depending on the original image size) and we expect their influence on the subsequent processing algorithms to be minor. Some categories of objects were less well handled. A number of large objects having uniform areas of density, such as pacemakers (1st case in Fig. 8), were not completely segmented and removed. As there were only a limited number of training examples with such large objects, we expect performance to increase with a larger database or dedicated

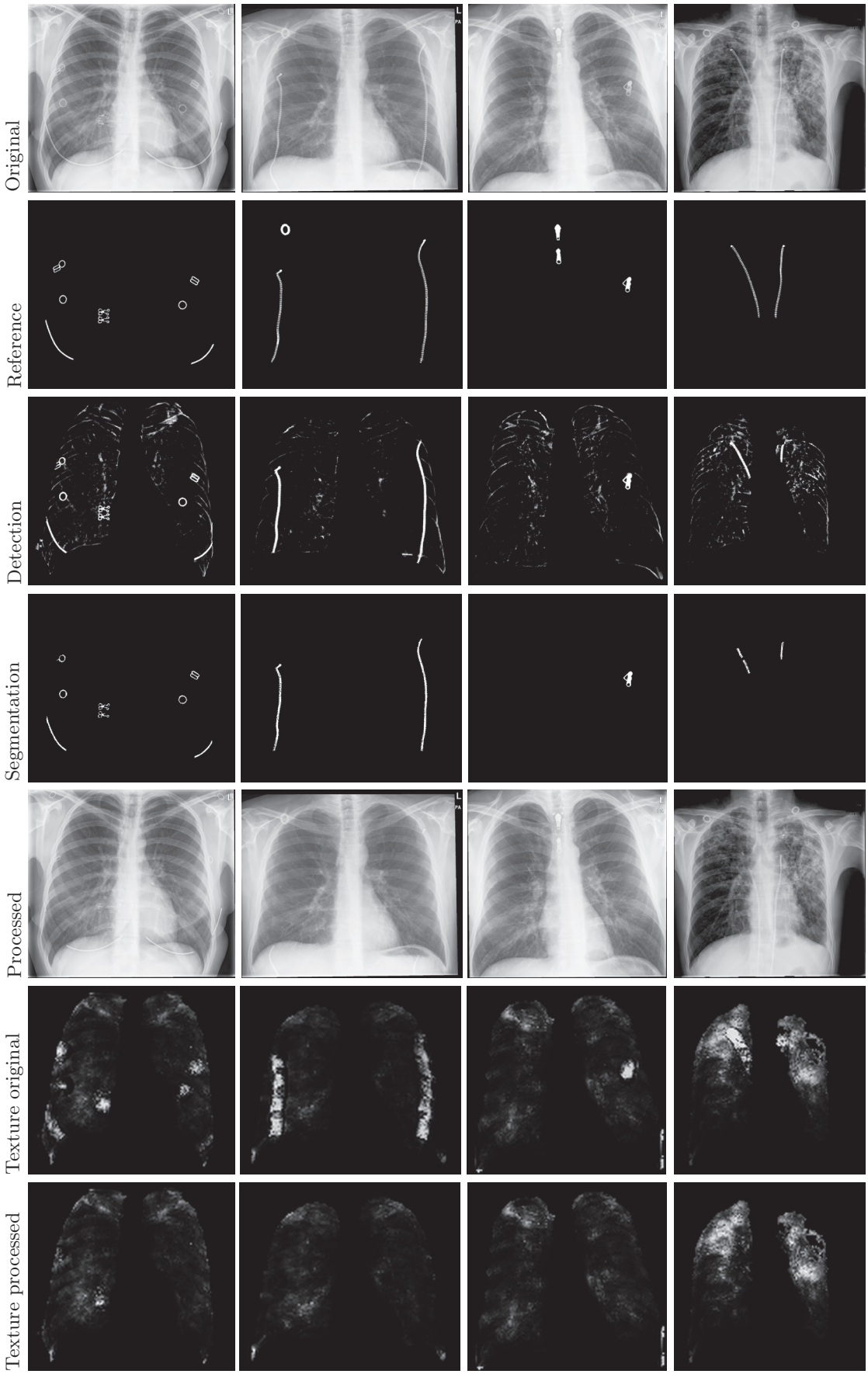


FIG. 7. Illustration of the output of the algorithm for three selected normal cases and one abnormal case (4th column) with foreign objects. See text for explanation.

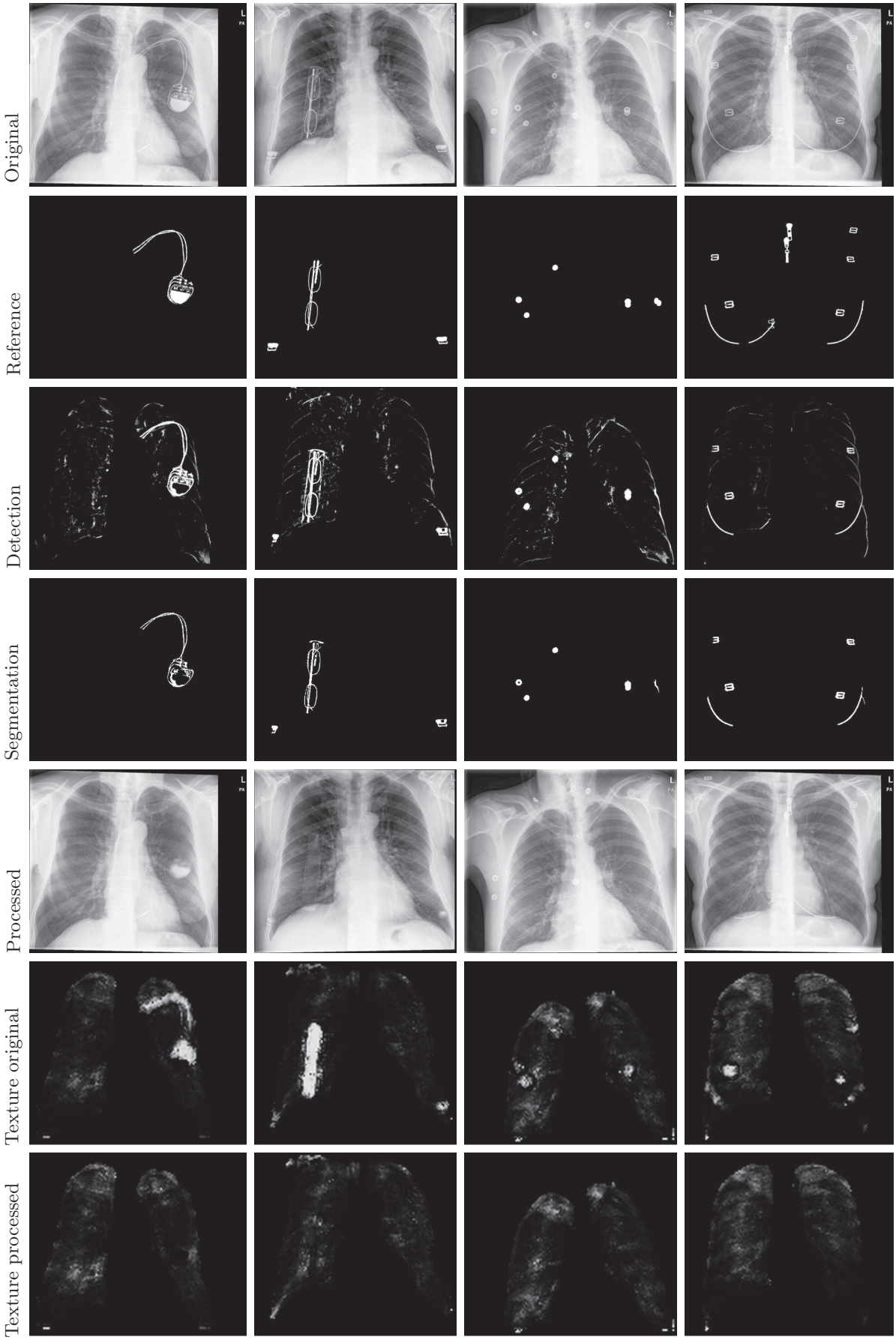


FIG. 8. Illustration of the output of the algorithm for four selected normal cases with foreign objects. See text for explanation.

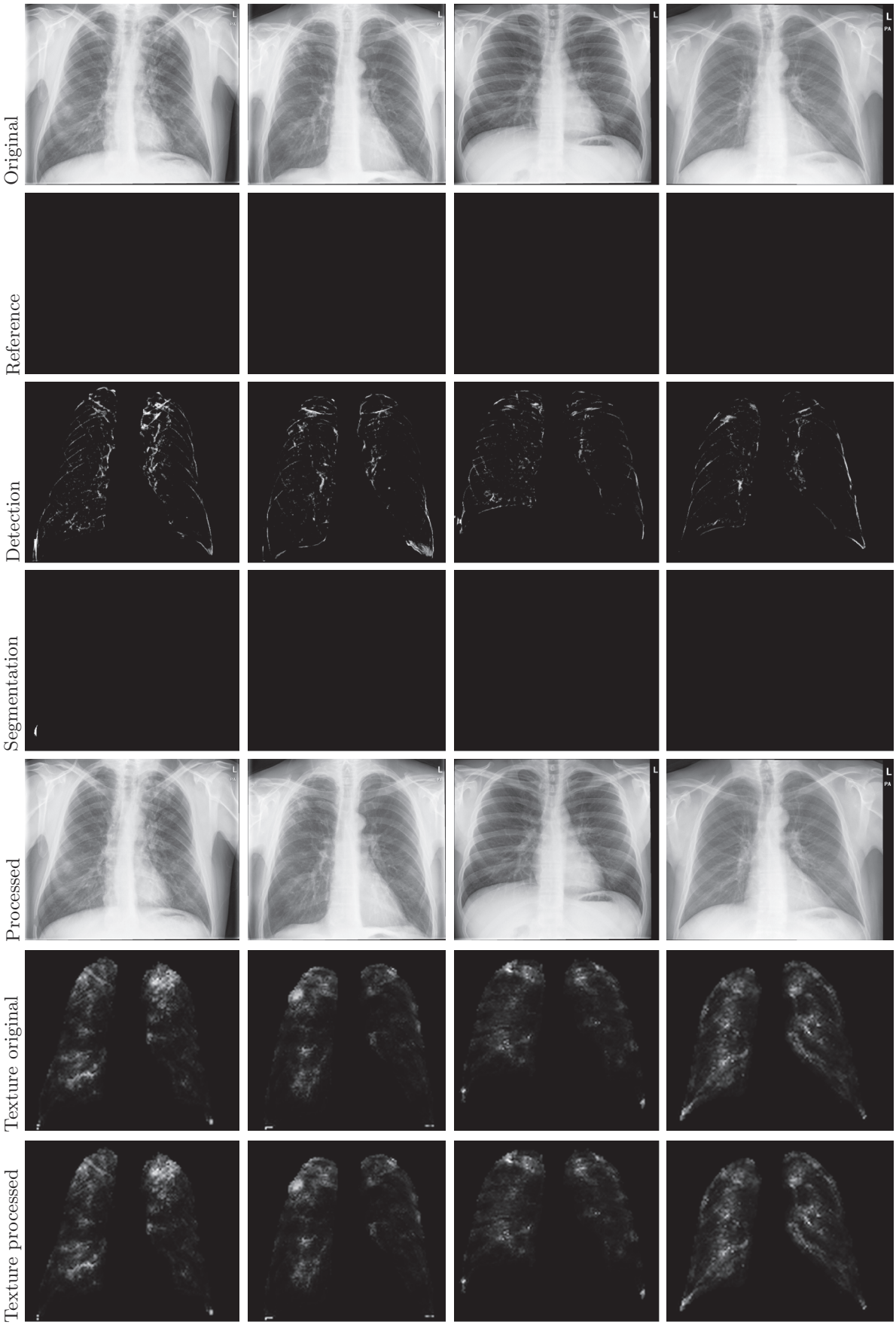


FIG. 9. Illustration of the output of the algorithm for four cases with no foreign objects. The first two cases contain pathology (1st case: abnormality in upper left lobe, 2nd case: abnormality in upper right lobe just below the clavicle), the last two cases contain no pathology.

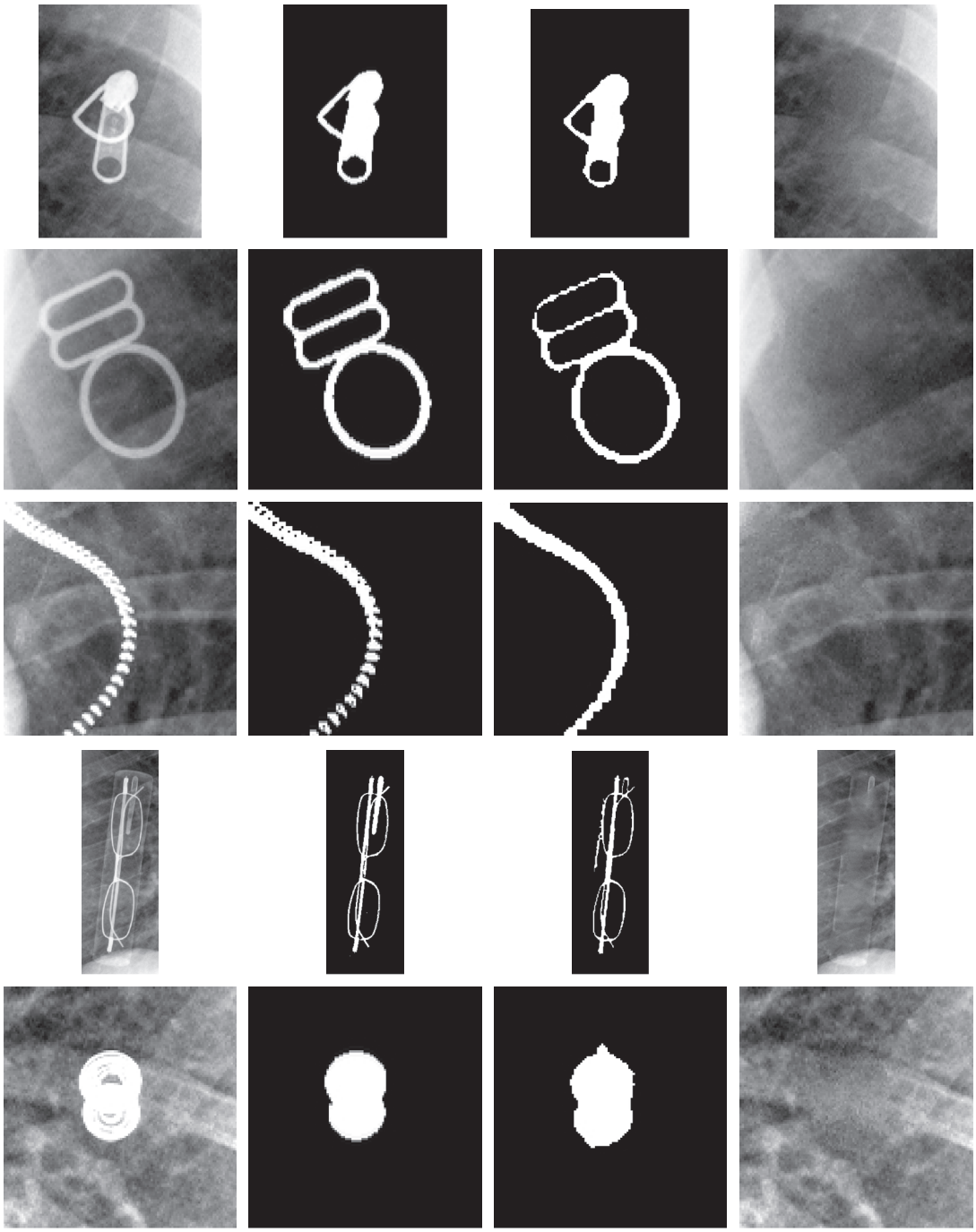


FIG. 10. Close-ups of several foreign objects. The columns show, respectively, the original image, the reference annotation, automatic segmentation and result of removal using texture synthesis. From top to bottom a zipper slider, brassier clip, zipper, glasses with case, and a button are shown.

set of training examples of pacemakers. Other types of foreign objects that were not or incompletely segmented were semi-opaque objects such as pens with plastic parts, hair, and certain types of necklaces. A good example is the case for glasses in the 2nd case in Fig. 8. The glasses were removed well, but the case itself remains visible. Also here we hypothesize that the reduced performance is partially caused by a limited number of similar training examples in the database. Semi-transparent objects are also more difficult to precisely delineate manually because of their lower contrast with the background. We expect though that these objects will have a relatively small disturbing effect on automated analysis as they occur less frequently and their appearance is less conspicuous.

At the probability cutoff used for segmentation, a false positive is detected in approximately one out of every four images (0.25 FP/image). Most of these false positives occur at the lung boundary, especially the interface of the lung with the chest wall (examples in the 4th case in Fig. 8 and 1st case in Fig. 9). We expect that the majority of this category

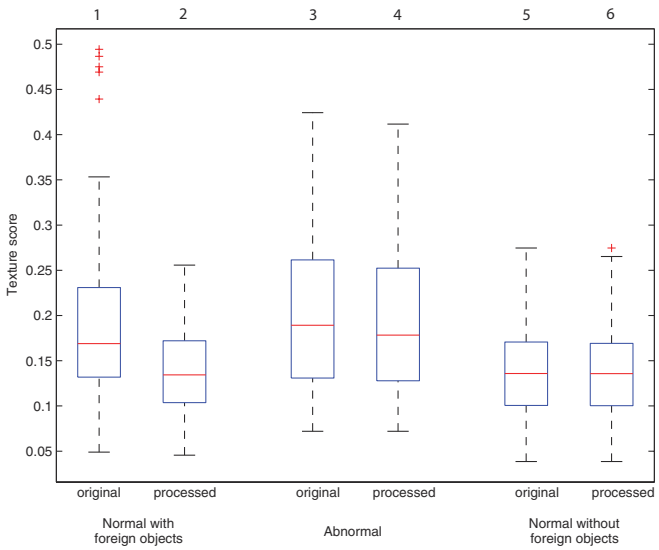


FIG. 11. Effect of restoration on texture scores for normal images containing foreign objects, normal images without foreign object and abnormal images. Boxplots of the texture score for the threeo groups of original and processed images.

of false positives would be easy to remove by adding extra position features to the classification, for example based on the distance to the lung border. False positives at the medial sides (near the heart and mediastinum) of the lung fields, or in the lung fields rarely occurred. Objects smaller than 120 pixels, corresponding to an effective diameter of 1.35–2.35 mm, were excluded in the FROC analysis and the texture analysis experiment. These objects are small compared to most foreign objects and we determined that also removing these objects from the image reduced texture scores only by 0.5%.

Detection and segmentation of foreign objects were limited to areas inside automatically detected unobscured lung fields. Depending on the location and appearance, objects close to the boundary of the lung fields can cause under- or over-segmentation of the lung fields. To also detect objects in over segmented lung fields, the observer providing manual segmentations was instructed to also segment objects close to the boundary of the lung fields. Examples of this can be seen in the 3rd case in Fig. 7 where a button is present close to, but outside of, the right upper lung and also in the 4th case of the same figure where the two zipper sliders at the top are outlined in the reference but outside the lung fields. As detection performance was only determined inside the lung fields this button was not counted as a false negative. Our main focus was to improve automatic analysis of the lung fields, therefore we limited the evaluation to this area. In some cases object close to, but outside, the automatic lung segmentation might disturb the texture analysis inside the lung fields. Although we expect the effect on the texture scores to be small, the algorithm could be easily extended to include all foreign objects in the radiograph.

In general, inpainting results were visually convincing. Poor results sometimes occurred when the object was not fully segmented. In these cases, the missed part will not be removed, but also the inpainting of the properly segmented parts can be distorted as there is a risk of selecting patches for inpainting from the missed part. In some cases, sharp edges and ridges, such as those caused by ribs, are not fully restored. The method used in this paper does not explicitly try to continue structural elements in the image before the addition of texture. It might be that other inpainting methods, such as those of Bertalmio *et al.*[12] would handle continuation of structures in a better way.

After restoration the texture scores of images containing foreign objects are similar to those of normal images. This indicates that the statistical properties of the affected areas, measured by the features used in the textural abnormality system, are similar to unaffected areas after removal. Using downstream image analysis algorithm performance, such as segmentation or CAD, is an indirect way to evaluate a detection and removal algorithm and visually the removal does not have to be perfect to provide results which improve subsequent analysis. This observation is also made by Lee *et al.*[8] who state that image artifacts after hair removal do not influence the segmentation of dermoscopic lesions. Evaluating a hair removal algorithm, Abbas *et al.*[17] used a similar approach as in this work, using segmentation performance and the effect on texture measures to determine the quality of the processed images.

Removal of detected foreign objects is based on a modified version of the texture synthesis algorithm described by Efros and Leung.[9] The modifications were aimed at increasing the speed of the algorithm, as the original algorithm is known to be extremely slow. Computation time was reduced from a few hours using the original algorithm to less than a minute with the modifications (C++ implementation on a single 3 Ghz core). Compared to previously published techniques for texture synthesis,[9,33] we have provided a method adapted to the intrinsic characteristics of medical images, particularly radiographs, with low computational time, which is a paramount feature for the analysis of large amount of images generated in screening setting. The speed increase can be beneficial when the algorithm is integrated into a CAD system that is used to provide feedback about image quality directly after acquisition of the radiograph. The full algorithm including feature calculation (15 s), pixel classification (4 min), segmentation (0 s), and inpainting (DullRazor 1 s, texture synthesis 30 s) takes approximately 10 min for an average case containing foreign objects. The majority of the computation time is spent on pixel classification using kNN. At a small loss of sensitivity, total computation times could be reduced to approximately 1 min when SVM or LDA is used instead.

The different steps of the algorithm, such as pixel classification, segmentation or texture synthesis, require a number of parameter values to be set. Some of them, such as the scales to compute features on or the minimum size of objects to retain, were based on the general characteristics of the foreign objects. Parameter settings of classifiers are difficult or impossible to determine *a priori* and were selected by optimization on the training set. Others, such as the segmentation threshold, depend on the performance characteristics of previous computations and were based on a trade-off between

over- and under segmentation. The choice of object removal algorithm and the free parameters of texture synthesis were not extensively optimized in our application. The selected settings resulted in successfully removal of foreign objects from images, with texture scores indistinguishable from unaffected images. Further optimization of some parameters might be possible to improve detection and removal of particularly different foreign objects.

An alternative approach to the presence of artifacts in medical images is to ignore affected areas. Such an approach potentially misses pathological areas. Many CAD systems work with feature extractors that have nonlocal support, which requires the size of the excluded area to be larger than the size of the object itself and would further increase the chance of false negatives. Therefore, we believe that a detection and removal algorithm is preferable.

In radiological screening settings numbers of abnormal images are often low, in the order of a few percent. In the tuberculosis screening program in London, which was the source of the data used in this paper, the number of abnormal chest radiographs which needed referral for abnormalities compatible with active tuberculosis was on the order of 1%. If an automatic system were to be used to select cases not needing referral a considerable improvement in efficiency, by a reduction in workload and costs, could be achieved. The percentage of images containing foreign objects in this database was 20%. If a significant proportion of this 20% were selected as cases needing referral the efficiency improvement would be much smaller.

The presented algorithm for detection and removal is general and can be applied to other objects in (chest) radiographs or to other modalities. In chest radiographs, medical foreign objects, such as catheters and tubes, are one of the most common abnormal findings, accounting for 64% of the total in a study by MacMahon *et al.*[44] In bedside radiographs, where these objects often occur, removing them can potentially improve the reading of these difficult low quality images by humans or automated systems.

VII. CONCLUSIONS

An automated method to detect, segment, and remove foreign objects on chest radiographs has been presented. The detection step is based on supervised pixel classification and evaluated using FROC analysis. The removal of the objects from the image is performed using texture synthesis inpainting. The effect of image restoration on false positive responses in a CAD system was determined in an experiment with a textural abnormality detection task.

We have found that high density foreign objects can be detected with high sensitivity with only a small number of false positives. The removal of the detected foreign objects from the image results in a reduction of false positive responses of a texture analysis system in normal images. This enables application of automated disease detection to improve efficiency of screening programs even with images of low quality due to the presence of foreign objects.

ACKNOWLEDGMENTS

This study was supported by the European and Developing Countries Clinical Trials Partnership (EDCTP) grant: the Evaluation of multiple novel and emerging technologies for TB diagnosis, in smear-negative and HIV-infected persons, in high burden countries (TB-NEAT) project. The authors would like to acknowledge the work of Jane Knight and Diana Taubman, the two reporting radiographers on the mobile X-ray unit in London who collected all of the CXRs.

[a]Author to whom correspondence should be addressed. Electronic mail: l.hogeweg@rad.umcn.nl; Telephone: +31 24 3653724; Fax: +31 24 3540866.

[1]M. Niemeijer, M. D. Abràmoff, and B. van Ginneken, "Image structure clustering for image quality verification of color retina images in diabetic retinopathy screening," Med. Image Anal. **10**, 888–898 (2006).

[2]S. Yoon, J. Feng, and A. K. Jain, "Altered fingerprints: Analysis and detection," IEEE Trans. Pattern Anal. Mach. Intell. **34**, 451–464 (2012).

[3]B. M. Keller, A. P. Reeves, M. D. Cham, C. I. Henschke, and D. F. Yankelevitz, "Semi-automated location identification of catheters in digital chest radiographs," Proc. SPIE **6514**, 65141O–65141O-9 (2007).

[4]K. Suzuki, H. Abe, H. MacMahon, and K. Doi, "Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN)," IEEE Trans. Med. Imaging **25**, 406–416 (2006).

[5]M. Loog and B. van Ginneken, "Bony structure suppression in chest radiographs," in *Computer Vision Approaches to Medical Image Analysis*, Lecture Notes in Computer Science Vol. 4241 (Springer Science+Business Media, Berlin, 2006).

[6]L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Physica D: Nonlinear Phenomena **60**, 259–268 (1992).

[7]D. Kuan, A. Sawchuk, T. Strand, and P. Chavel, "Adaptive restoration of images with speckle," IEEE Trans. Acoust., Speech, Signal Process. **35**, 373–383 (1987).

[8]T. K. Lee, V. Ng, R. Gallagher, A. Coldman, and D. McLean, "Dullrazor: A software approach to hair removal from images," Comput. Biol. Med. **27**, 533–543 (1997).

[9]A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proceedings of Seventh IEEE International Conference on Computer Vision, 1999* (IEEE, New York, 1999), Vol. 2, pp. 1033–1038.

[10]M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '00* (Association for Computing Machinery, New York, 2000), pp. 417–424.

[11]G. Emile-Male, *The Restorer's Handbook of Easel Painting*, 1st ed. (Van Nostrand Reinhold, New York, 1976).

[12]M. Bertalmio, L. Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," IEEE Trans. Image Process. **12**, 882–889 (2003).

[13]A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," IEEE Trans. Image Process. **13**, 1200–1212 (2004).

[14]H. Zhou, M. Chen, R. Gass, J. M. Rehg, L. Ferris, J. Ho, and L. Drogowski, "Feature-preserving artifact removal from dermoscopy images," Proc SPIE **6914**, 69141B (2008).

[15]P. Wighton, T. K. Lee, and M. S. Atkins, "Dermascopic hair disocclusion using inpainting," Proc. SPIE **6914**, 691427 (2008).

[16]F. Bornemann and T. März, "Fast image inpainting based on coherence transport," J. Math. Imaging Vision **28**, 259–278 (2007).

[17]Q. Abbas, I. F. Garcia, M. E. Celebi, and W. Ahmad, "A feature-preserving hair removal algorithm for dermoscopy images," Skin Res. Technol. **19**(1), 1–10 (2011).

[18]A. Story, S. Murad, W. Roberts, M. Verheyen, A. C. Hayward, and London Tuberculosis Nurses Network, "Tuberculosis in London: The importance of homelessness, problem drug use, and prison," Thorax **62**, 667–671 (2007).

[19]A. M. R. Schilham, B. van Ginneken, and M. Loog, "A computer-aided diagnosis system for detection of lung nodules in chest radiographs with

an evaluation on a public database," Med. Image Anal. **10**, 247–258 (2006).
[20]B. van Ginneken, M. B. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database," Med. Image Anal. **10**, 19–40 (2006).
[21]A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," IEEE Trans. Pattern Anal. Mach. Intell. **22**, 4–37 (2000).
[22]R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. (Wiley, New York, 2001).
[23]L. M. J. Florack, B. M. ter Haar Romeny, M. A. Viergever, and J. J. Koenderink, "The Gaussian scale-space paradigm and the multiscale local jet," Int. J. Comput. Vis. **18**, 61–75 (1996).
[24]R. Deriche, "Fast algorithms for low-level vision," IEEE Trans. Pattern Anal. Mach. Intell. **12**, 78–87 (1990).
[25]A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science Vol. 1496 (Springer-Verlag, Berlin, 1998), pp. 130–137.
[26]S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu, "An optimal algorithm for approximate nearest neighbor searching in fixed dimensions," J. ACM **45**, 891–923 (1998).
[27]J. Friedman, T. Hastie, and R. Tibshirani, "Special invited paper. Additive logistic regression: A statistical view of boosting," Ann. Stat. **28**, 337–374 (2000).
[28]A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing visual features for multiclass and multiview object detection," IEEE Trans. Pattern Anal. Mach. Intell. **29**, 854–869 (2007).
[29]R. Lienhart, A. Kuranov, and V. Pisarevsky, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection," in *Pattern Recognition*, Lecture Notes in Computer Science Vol. 2781 (Springer-Verlag, Berlin, 2003), pp. 297–304.
[30]A. K. Jain and S. Z. Li, *Handbook of Face Recognition* (Springer-Verlag, New York, 2005).
[31]C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn. **20**, 273–97 (1995).
[32]C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM Trans. Intell. Syst. Technol. **2**(3), 27:1–27:27 (2011).
[33]L. Y. Wei and M. Levoy, "Fast texture synthesis using tree-structured vector quantization," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '00* (2000).
[34]C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," see http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.
[35]P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," Pattern Recogn. Lett. **15**, 1119–1125 (1994).
[36]B. Efron, "Bootstrap methods: Another look at the jackknife," Ann. Stat. **7**, 1–26 (1979).
[37]P. C. Bunch, J. F. Hamilton, G. K. Sanderson, and A. H. Simmons, "A free response approach to the measurement and characterization of radiographic-observer performance," J. Appl. Photogr. Eng. **4**, 166–172 (1978).
[38]R. Zwiggelaar, C. Taylor, and C. Rubin, "Detection of the central mass of spiculated lesions: Signature normalisation and model data aspects," in *Information Processing in Medical Imaging*, Lecture Notes in Computer Science Vol. 1613 (Springer-Verlag, Berlin, 1999).
[39]A. Oliver, J. Freixenet, J. Martí, E. Pérez, J. Pont, E. R. E. Denton, and R. Zwiggelaar, "A review of automatic mass detection and segmentation in mammographic images," Med. Image Anal. **14**, 87–110 (2010).
[40]Y. Arzhaeva, M. Prokop, D. M. J. Tax, P. A. de Jong, C. M. Schaefer-Prokop, and B. van Ginneken, "Computer-aided detection of interstitial abnormalities in chest radiographs using a reference standard based on computed tomography," Med. Phys. **34**, 4798–4809 (2007).
[41]L. Hogeweg, C. Mol, P. A. de Jong, R. Dawson, H. Ayles, and B. van Ginneken, "Fusion of local and global detection systems to detect tuberculosis in chest radiographs," in *Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science Vol. 6363 (Springer-Verlag, Berlin, 2010).
[42]M. Loog and B. van Ginneken, "Static posterior probability fusion for signal detection: applications in the detection of interstitial diseases in chest radiographs," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004* (2004), pp. 644–647.
[43]R. McGill, J. W. Tukey, and W. A. Larsen, "Variations of box plots," Am. Stat. **32**, 12–16 (1978).
[44]H. MacMahon, K. J. M. Liu, S. M. Montner, and K. Doi, "The nature and subtlety of abnormal findings in chest radiographs," Med. Phys. **18**, 206–210 (1991).