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

Laurens Hogeweg, 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. © 2013 American Association of Physicists in Medicine. [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 and fingerprints.

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.
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 (without the aim of actually removing the catheters) and the removal of anatomical structures such as ribs and clavicles by suppressing them. 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. 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. 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 who employed nonparametric sampling. Bertalmio et al. 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. 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. used anisotropic diffusion to propagate linear structures along isophotes. In an improved version texture was also added. Criminisi et al. 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. used automatic line extraction followed by inpainting to remove the objects. Wighton et al. 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. In a more recent comparison it was found that fast marching inpainting performed better than linear interpolation, nonlinear diffusion and exemplar-based inpainting in hair removal judged by segmentation performance and texture analysis. In this work, we employ a modified version of the texture synthesis method described by Efros and Leung 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,
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. 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 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

---

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

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

**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.
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.\textsuperscript{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 \textit{et al.}\textsuperscript{20} Features were calculated at images of order 0, 1, 2 ($L_x$, $L_y$, $L_{xx}$, $L_{xy}$, $L_{yy}$), 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\textsuperscript{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 ($L_x$, $L_y$) 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.\textsuperscript{25} Considering the two eigenvalues of the Hessian matrix $\lambda_1$, $\lambda_2$ with $|\lambda_1| > |\lambda_2|$, two measures were derived: (1) $\sqrt{(\lambda_1^2 - \lambda_2^2)}$ to extract the linearness 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. Detection of foreign objects

III.B.1. Pixel classification (PC)

In this methodology, the segmentation problem is recast into a pattern classification task.\textsuperscript{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,\textsuperscript{23} the output of Gaussian derivative filtered images of order 0, 1, 2 ($L_x$, $L_y$, $L_{xx}$, $L_{xy}$, $L_{yy}$), 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\textsuperscript{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 ($L_x$, $L_y$) 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.\textsuperscript{25} Considering the two eigenvalues of the Hessian matrix $\lambda_1$, $\lambda_2$ with $|\lambda_1| > |\lambda_2|$, two measures were derived: (1) $\sqrt{(\lambda_1^2 - \lambda_2^2)}$ to extract the linearness 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 \((x_1, y_1), ..., (x_N, y_N)\) where \(x_i \in X, y_i \in Y = \{-1, +1\}\)
2. Start with \(H(x_i) = 0\) and weights \(w_i = 1/N, i = 1, ..., N\)
3. Repeat for \(m = 1, ..., M\)
   \[\text{(a) Find the optimal weak classifier } h_m \text{ for } (X, Y) \text{ and current weights } w_i\]
   \[\text{(b) Update strong classifier } H(x) \leftarrow H(x) + h_m(x)\]
   \[\text{(c) Update weights for examples } w_i \leftarrow w_i e^{-y_i h_m(x_i)} \text{ for } i = 1, ..., N\]

where \(k_i\) 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 \(\epsilon\), 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. \(\epsilon\) 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.\(^{25}\) 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

\[J_m = \sum_{i=1}^{N} w_i (y_i - h_m(x_i))^2,\quad (4)\]

where \(w_i = e^{-y_i H(x_i)}\) 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 \(h_m(x_i) = a [x_i > \theta] + b [x_i < \theta]\), where \(f\) is the the feature number and \(\delta\) 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 \(H(x) = \log P(y|x)/P(-y|x)\), \(y \in [-1, +1]\). The posterior probability for \(y = 1\) given \(x\) is estimated using a sigmoid function as follows:

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

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:

$$\begin{align*}
\text{minimize} & \quad \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \\
\text{subject to} & \quad y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i \\
& \quad \xi_i \geq 0,
\end{align*}$$

where $w$ 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 $\xi$, which are introduced to allow for solutions when a hyperplane splitting the classes does not exist. The function $\phi$ maps the feature vectors into a higher dimensional space where a hyperplane may be easier to find. $\phi$ is related to a kernel function $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$. Different kernel functions can be used. Setting $K(x_i, x_j) = x_i^T x_j$ gives a linear kernel. The Gaussian radial basis function $K(x_i, x_j) = \exp(-\gamma ||x_i - x_j||^2)$, $\gamma > 0$ is often used as a nonlinear kernel and introduces an extra parameter $\gamma$. SVM does not provide posterior probabilities $P(y = 1|x)$ 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. Training times for SVM can be considerable as $C$ and $\gamma$ 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 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, a fast version of kNN search (inspired by the use of tree structured vector quantization in Wei and Levoy). 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 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 \( k \text{NN} \) 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 \( \gamma \) were optimized in a grid search by cross validation on the training set, according to the recommendations in Hsu et al.\(^{36} \) The optimal values of \( C \) and \( \gamma \) 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 \( A_\text{ROC} \). Differences between \( A_\text{ROC} \) 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 \( \Omega = \text{TP}/(\text{TP} + \text{FN}) \). 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, \( p_\text{t} \) 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 \( p_\text{t} \) 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 \( p_\text{t} \). 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 \( p_\text{t} = 1.0 \). At each level of \( p_\text{t} \), FP objects are identified in the segmentation. For each FP object it is determined whether it overlaps with an object detected at a higher \( p_\text{t} \). If it was previously detected it is ignored, otherwise the object is recorded as FP and assigned a score equal to the current \( p_\text{t} \). In a similar way the TP and FN objects are assigned a score based on the level of \( p_\text{t} \) where they first appeared. In the reported results, the step size for \( p_\text{t} \) 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 tran-

**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 \( A_\text{ROC} \) were found between LDA, the best performing classifier, and the other classifiers. At high specificities (0.99–1.0), the operating region for segmentation, \( k \text{NN} \) and SVM are the best classifiers [Fig. 5(b)]. Nearest mean classification had the lowest performance but still reached a high overall \( A_\text{ROC} \) value. Feature selection did not lead to improved sensitivity at the level of 0.995 specificity, except for nearest mean classification, which
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 $\lambda_1$ 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>
<thead>
<tr>
<th>No.</th>
<th>Individual performance</th>
<th>Sensitivity at 0.995 specificity</th>
<th>Selection order in SFS</th>
<th>Sensitivity at 0.995 specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\lambda_1 \ (\sigma = 1 \ \text{pixels})$</td>
<td>0.356</td>
<td>$\lambda_1 \ (\sigma = 1 \ \text{pixels})$</td>
<td>0.356</td>
</tr>
<tr>
<td>2</td>
<td>$L_{yy} \ (\sigma = 2 \ \text{pixels})$</td>
<td>0.316</td>
<td>$L_{xx} \ (\sigma = 4 \ \text{pixels})$</td>
<td>0.516</td>
</tr>
<tr>
<td>3</td>
<td>$L_{yy} \ (\sigma = 4 \ \text{pixels})$</td>
<td>0.288</td>
<td>$L_{yy} \ (\sigma = 4 \ \text{pixels})$</td>
<td>0.643</td>
</tr>
<tr>
<td>4</td>
<td>$\lambda_1 \ (\sigma = 2 \ \text{pixels})$</td>
<td>0.285</td>
<td>$L_{y} \ (\sigma = 64 \ \text{pixels})$</td>
<td>0.674</td>
</tr>
<tr>
<td>5</td>
<td>$L_{yy} \ (\sigma = 1 \ \text{pixels})$</td>
<td>0.271</td>
<td>$L_{y} \ (\sigma = 1 \ \text{pixels})$</td>
<td>0.743</td>
</tr>
</tbody>
</table>

Medical Physics, Vol. 40, No. 7, July 2013
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.
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. 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.
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. 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. who state that image artifacts after hair removal do not influence the segmentation of dermoscopic lesions. Evaluating a hair removal algorithm, Abbas et al. 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. 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, 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.\textsuperscript{4}\textsuperscript{4} 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.

\section{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 temporal 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.

\begin{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.
\end{acknowledgments}

\textsuperscript{4}Author to whom correspondence should be addressed. Electronic mail: l.hogeweg@rad.umcn.nl; Telephone: +31 24 3653724; Fax: +31 24 3540866.
\textsuperscript{23}A. M. R. Schilham, B. van Ginneken, and M. Loog, “A computer-aided diagnosis system for detection of lung nodules in chest radiographs with


