Automated Detection of Uninformative Frames in Pulmonary Optical Endomicroscopy

Antonios Perperidis*, Ahsan Akram, Yoann Altmann, Paul McCool, Jody Westerfeld,
David Wilson, Kevin Dhaliwal, and Stephen McLaughlin

Abstract—*Significance:* Optical endomicroscopy (OEM) is a novel real-time imaging technology that provides endoscopic images at a microscopic level. The nature of OEM data, as acquired in clinical use, gives rise to the presence of uninformative frames (i.e., pure-noise and motion-artefacts). Uninformative frames can comprise a considerable proportion (up to >25%) of a dataset, increasing the resources required for analyzing the data (both manually and automatically), as well as diluting the results of any automated quantification analysis. *Objective:* There is, therefore, a need to automatically detect and remove as many of these uninformative frames as possible while keeping frames with structural information intact. *Methods:* This paper employs Gray Level Cooccurrence Matrix texture measures and detection theory to identify and remove such frames. The detection of pure-noise and motion-artefacts frames is treated as two independent problems. *Results:* Pulmonary OEM frame sequences of the distal lung are employed for the development and assessment of the approach. The proposed approach identifies and removes uninformative frames with a sensitivity of 93% and a specificity of 92.6%. *Conclusion:* The detection algorithm is accurate and robust in pulmonary OEM frame sequences. Conditional to appropriate model refinement, the algorithms can become applicable in other organs.

Index Terms—Distal lung imaging, fibered confocal fluorescent microscopy (FCFM), frames detection, image analysis, optical endomicroscopy (OEM), texture analysis.

I. INTRODUCTION

OPTICAL endomicroscopy (OEM) is an emerging imaging tool used both clinically and preclinically [1]. Fibered confocal fluorescent microscopy (FCFM), also referred to as probe-based confocal laser endomicroscopy, is the most widely used platform and the only fiber-based endomicroscopic methodology approved for clinical use. The technology employs a proximal laser scanning unit linked to an interface with a flexible multicore fiber. This fiber is passed through the working channel of endoscopes enabling microscopic imaging at the distal end of the fiber. In pulmonary OEM, the abundance of elastin and collagen enables structural imaging through the generation of autofluorescence with a 488-nm laser excitation. The lateral diameter of the fiber used in lung applications is 1.4 mm. This miniaturization enables the exploration of the distal pulmonary tract [2] as well as the assessment of the respiratory bronchioles and alveolar gas exchanging units of the distal lung [3]. OEM has been used clinically in the lung for the detection of lung cancer [4], [5] and has been used to assess the distal lung [6], [7] including the imaging of parenchymal lung diseases [8]. Furthermore, OEM has been used in other organs such as the urological tract [9]. The largest OEM application remains in imaging of possible cancerous lesions in the gastrointestinal tract [10], [11]. The commercially available FCFM platform images at 12 frames/s and clinical and preclinical OEM procedures often last minutes, generating thousands of frames; hence, making their manual (post*vivo*) analysis a very labor intensive process.

The nature of OEM data acquisition results in image sequences that form a long continuous scene. Within these sequences, there are frames that contain only pure-noise [see Fig. 1(a)], mostly due to the lack of contact of the fiber with a fluorescent target or due to bio fouling of the tip of the fiber. Similarly, there are frame sequences where the spatial movement is very large when compared to the temporal rate of acquisition. This results in motion artefacts, expressed as either deformed anatomical structures [see Fig. 2(a)–(c)], or spatial discontinuity for temporally adjacent frames [see Fig. 2(d)–(f)]. Such frames contain a little information of value and are, therefore, referred to as "uninformative frames." Indeed, uninformative frames comprise a substantial proportion of the dataset, depending on the motion of the imaging target as well as also the operator manipulating the fiber. In pulmonary OEM, significant movement artefacts occur due to the movement of the fiber in the distal lung from both the respiratory effort of the patient and also of the fiber traversing bronchopulmonary segments of the lung. In our experience with lung OEM data, uninformative frames may comprise in excess of 25% of the acquired frames. The presence of uninformative frames: 1) prolongs the offline manual assessment of the data, 2) increases computational resources, and 3) dilutes the results of any postprocessing algorithm intended to analyze and quantify the images. There is, therefore,
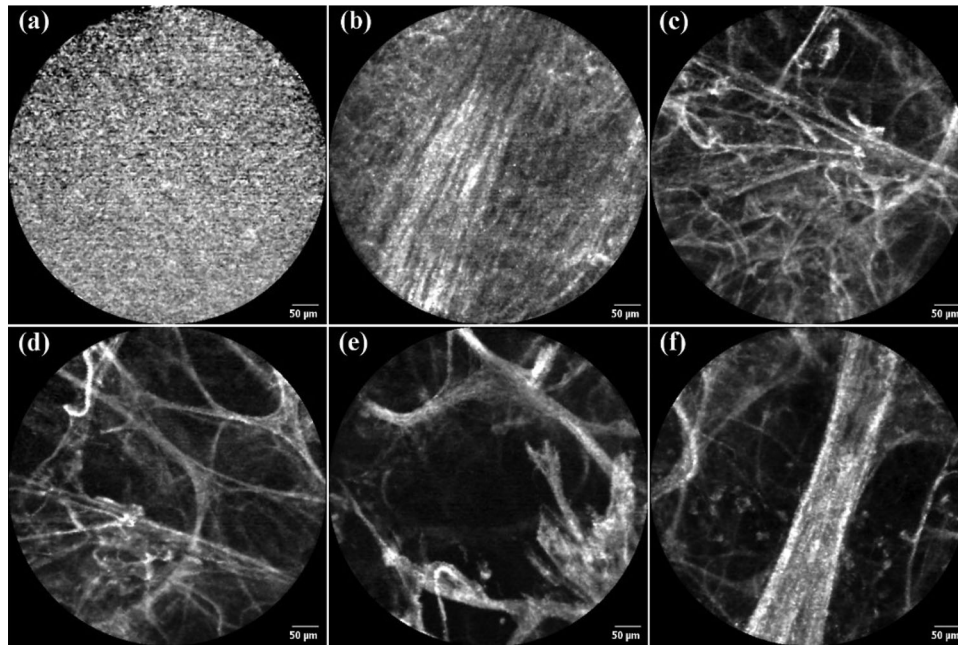
Fig. 1. Representative examples of frames containing (a) pure noise, (b) low contrast and mostly linear bronchus strands, (c), (d) normal elastin strands, pathological and healthy, respectively, (e) larger elastin strands, and (f) blood vessels.

a need for an automated approach to accurately and robustly detect and remove such uninformative frames as the first line of any automated or manual image analysis step.

There has been considerable research in the suppression of the noise [12] as well as for the detection of motion artefacts for a range of imaging modalities, including, but not limited to, aerial images [13], microscopy [14], medical images [15], as well as other digital photography images [16]–[18]. However, most such studies focus on the detection of motion-blurred regions within a frame with the intention to compensate for it through some image enhancement algorithm. Such techniques, while potentially very effective for their specific application, cannot be easily employed to detect uninformative frames in OEM data. Analyzing the spatiotemporal characteristics of the sequences is required. A large number of studies performing such analysis of the spatiotemporal characteristics of video sequences concentrate on detecting shot transitions and grouping frames into scenes [19]–[22]. Once again, this is not applicable in OEM data due to the continuous acquisition resulting in a continuous imaging sequence with uninformative frames embedded within it. Other endoscopic imaging techniques can generate analogous frame sequences to OEM when navigating along the bronchus, or the gastrointestinal tract. There is, therefore, considerable interest in the spatiotemporal analysis of endoscopic data, including, but not limited to, laparoscopy [23], colonoscopy [24], wireless capsule endoscopy [25]–[28], and larynx endoscopy [29]. The main focus of all these studies was the identification of one or more key frames within the main frame sequence to aid the diagnostic process or some further postprocessing technique. A recent study [30] has developed a fully automated approach for the selection of a representative frame from a short endomicroscopy frame sequence, enabling a real-time quantitative image analysis at the point-of-care. The approach generated

very promising results for short oral and esophageal image sequences. However, none of the aforementioned studies address the problem of identifying and uninformative frames from OEM frame sequences.

This paper presents a novel approach for detecting and "removing" uninformative frames from OEM frames sequences. The algorithm was developed and assessed on frame sequences from the distal lung of patients with suspected lung cancer. However, with the appropriate adjustments, the algorithm can potentially be effective in removing uninformative frames from sequences acquired on 1) other organ systems, such as the gastrointestinal tract and the urinary tract, as well as 2) any other fiber-based imaging platform. The rest of this paper is organized as follows. Section II describes the material (data) utilized in this study. Section III describes the detection algorithms for pure-noise and motion-artefacts independently. Section IV describes the data analysis used to train and test the detection algorithm, and Section V displays the relevant results. Finally, the proposed methods and corresponding results are discussed in Section VI.

II. DATA

Eighty-three OEM image sequences of the distal lung were used during the development and testing of the proposed algorithm. All data were obtained as part of a database (of 126 subjects) during the routine care of patients undergoing investigation for an indeterminate pulmonary nodule (< 30 mm) at the Columbus Lung Institute, Indiana, USA. The study was approved by the Western Institutional Review Board. All procedures were undertaken by a single expert operator using standard bronchoscopy, with the aid of a superDimension Navigation System (Covidien Inc., MN, USA) and imaging with 488-nm Cellvizio using a 1.4-mm lateral diameter Alveoflex fiber (Mauna Kea Technologies, Paris, France). All image
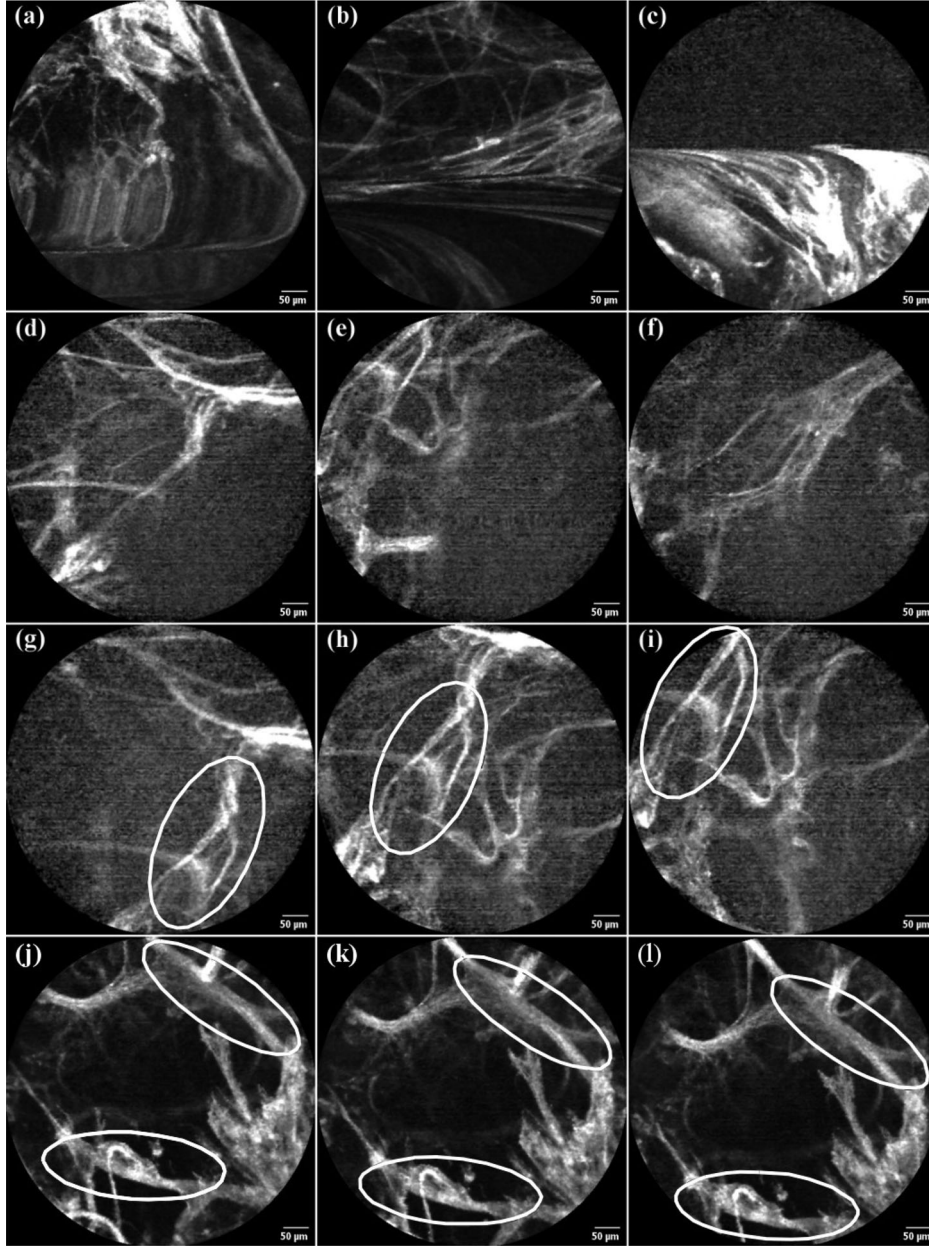
Fig. 2. Representative examples of frames sequences containing (a)–(c) motion artefacts (deformed structures), (d)—(f) motion artefacts (spatial discontinuity in temporally adjacent frames), (g)—(i) large movements, and (j)–(l) normal/modest movements. The circular regions highlight structures that demonstrate the scale of the movement.

sequences were stored in the proprietary .mkt format and read as 16-bit binary files for processing in MATLAB (MathWorks, Inc., MA, USA). Some subjects ($n = 43$) were rejected due to 1) short duration of sequences (i.e., video $<$ 10 frames), 2) corrupted data (i.e., file not readable, misaligned fiber, or out of focus images), or 3) lack of distal lung images (i.e., solely imaging the bronchus). No other subjective criteria (such as image quality) that could potentially bias the proposed algorithm were used during the video selection process.

III. METHODOLOGY

This section describes the methodology used to detect pure-noise and motion-artefacts frames. These were handled as two independent problems, both utilizing image-derived texture metrics.

Let $I(x, y, t)$ be a gray scale image sequence, with $x \in [1, N]$, $y \in [1, M]$, and $t \in [1, K]$ indicating the pixel location (x—column and y—row) and the frame number, respectively. The Gray-Level Cooccurrence Matrix (GLCM) [31] G_t for frame $I_t = I(x, y, t)_{(x,y) \in [1,N] \times [1,N]}$ was defined as an $L \times L$ matrix

$$G_t^{\Delta x, \Delta y}(i, j)$$
$$= \sum_{p=1}^{N} \sum_{q=1}^{M} \begin{cases} 1, & \text{if } I(p, q, t) = i \text{ and } I(p + \Delta x, q + \Delta y, t) = j \\ 0, & \text{otherwise} \end{cases}$$
$$(1)$$

where L was the number of gray levels within the image (16 bit), i and j were intensity levels, p and q were the spatial positions in the image I_t, and Δx and Δy were the spatial offsets (in number of pixels) utilized to estimate the GLCM G_t. In order to achieve rotational invariance of the relevant texture measures, G_t was estimated as the mean GLCM for four different offset pairs $\{(1,0),(1,1),(0,1),(-1,1)\}$, corresponding to a single pixel offset at directions ($0°$, $45°$, $90°$, and $135°$). G_t was then normalized to denote the probability of each pixel pair being present in frame I_t

$$G_t^{\mathrm{norm}} = \frac{1}{n} G_t \tag{2}$$

where n was the sum of all the elements of the matrix G_t. Related texture metrics were derived as [31]:

$$\text{Contrast}\,(t) = V_1(t) = \sum_{i=1}^{L} \sum_{j=1}^{L} (i-j)^2\, p_{\mathrm{ij}} \tag{3}$$

$$\text{Energy}\,(t) = V_2\,(t) = 1 - \sum_{i=1}^{L} \sum_{j=1}^{L} p_{\mathrm{ij}}^2 \tag{4}$$

$$\text{Homogeneity}\,(t) = V_3\,(t) = 1 - \sum_{i=1}^{L} \sum_{j=1}^{L} \frac{p_{\mathrm{ij}}}{1 - |i-j|} \tag{5}$$

$$\text{Entropy}(t) = V_4(t) = - \sum_{i=1}^{L} \sum_{j=1}^{L} p_{\mathrm{ij}} \log_2(p_{\mathrm{ij}}) \tag{6}$$

$$\text{Maximum Probability}\,(t) = V_5\,(t) = 1 - \max_{i,j}(p_{\mathrm{ij}}) \tag{7}$$

where t was the frame number and $p_{\mathrm{ij}} = G_t^{\mathrm{norm}}(i, j)$. In addition to the aforementioned GLCM properties, global image characteristics, such as frame intensity mean ($V_6(t)$), and standard deviation ($V_7(t)$), were also employed. All texture metrics were estimated in a way such that frames containing noise (or very faint features) demonstrated low (nearly zero) values [see Fig. 1(a) and (b)], while more pronounced features, such as elastin strands and blood vessels [see Fig. 1(c)–(f)] within the alveoli space, demonstrated higher (closer to 1) values. Since the GLCMs need to be estimated in rectangular regions only, the largest square region within the circular field of view (FOV) of the OEM frame sequences was used as $I(x, y, t)$ throughout this study. The remaining four segments (each 9% of the overall circular FOV) were not included in $I(x, y, t)$, and, consequently, in the GLCM estimation and the subsequent frame detection. This decision was based in the assumption that, if the central square region of a frame was identified as pure-noise or a motion-artefact, a small structure in any of the four excluded subsections is not enough to reinstate the frame as an informative frame.

Each texture vector was normalized to the [0, 1] range using

$$V_i'(t) = \frac{V_i(t) - \min(V_i)}{\max(V_i) - \min(V_i)} \tag{8}$$

where $i \in [1, 7]$ and $t \in [1, K]$ represented the frame number. Finally, a 7-D feature space was, therefore, defined as

$$X = [V_i']. \tag{9}$$

A. Detection of Pure-Noise Frames

1) Reducing Dimensionality: Principal component analysis (PCA) was employed to reduce the dimensionality of the 7-D feature space. The $K \times 7$ matrix $Y = (\mathrm{PC1}, \mathrm{PC2}, \dots, \mathrm{PC7})$ was defined as

$$Y = A(X - (\lambda * m_X)) \tag{10}$$

incorporating the projection of the seven feature vectors (X) in the relevant principal component space. The vector $m_X = \frac{1}{K} \sum_{t=1}^{K} X(t, i), i \in [1, 7]$ contained the mean value of each of the seven parameters and the vector λ was an $1 \times K$ unit vector, while the rows of A were the eigenvectors (i.e., the direction of the principal components) of C_X, the 7×7 sample covariance matrix of X (normalized to unit length)

$$C_X = \frac{1}{K-1} \sum_{t=1}^{K} (X(t) - m_X)(X(t) - m_X)^T. \tag{11}$$

Although the whole matrix Y could be used for the detection of pure-noise frames, the first principal component (PC1) was found to contain sufficient information for the detection of pure-noise frames. As a consequence, only PC1 (i.e., a single parameter per frame) has been considered for pure-noise frame detection.

2) Gaussian Mixture (GM) Model: An experienced investigator performed a thorough visual inspection on a subset of the available OEM data, aiming to identify any image texture subgroups that can justly represent the underlying anatomical information. The inspection of the OEM data highlighted four different texture categories (see Fig. 1): 1) pure-noise frames [see Fig. 1(a)], mostly containing no anatomical information, 2) subtle feature frames [see Fig. 1(b)], mostly containing linear bronchus strands or very low contrast elastin strands, 3) normal frames [see Fig. 1(c)-(d)], containing both pathological [see Fig. 1(c)] and healthy [see Fig. 1(d)] elastin strands, and 4) vibrant frames [see Fig. 1(e) and (f)], containing very well defined features, such as larger elastin strands and blood vessels. The boundaries of these four categories were not distinct.

Fig. 3(a) provides a representative histogram example $H1$ derived from the PC1 of a lung OEM image sequence. A GM model was employed to represent the underlying texture information contained in PC1. More precisely, following the four texture categories identified through the aforementioned manual visual inspection of the OEM data, the following GM model composed of 4 Gaussian distributions was considered

$$\mathrm{GM} = \sum_{i=1}^{4} P_i \mathcal{N}_i \tag{12}$$

where parameters P_i provided the weight (also referred to as proportion or probability $\sum_{i=1}^{4} P_i = 1$) of the ith Gaussian
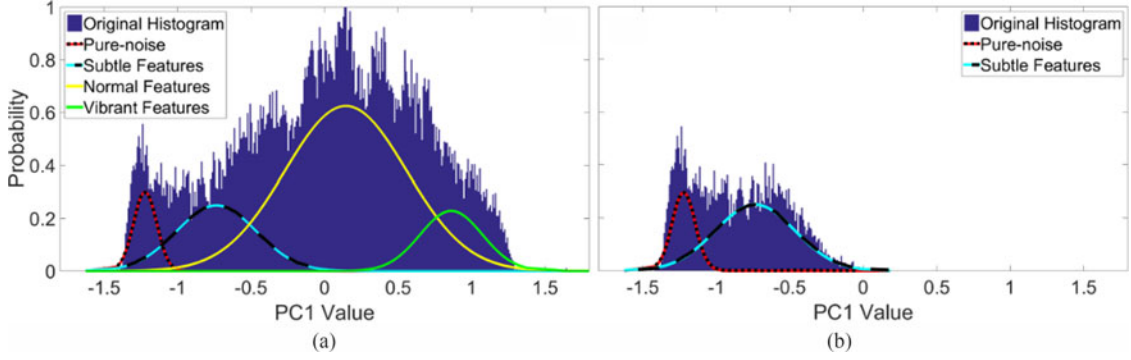
Fig. 3. (a) Histogram corresponding to the PC1 from 72 frame sequences concatenated as a single dataset along with the corresponding 4-GM model. (b) Refined histogram along with the corresponding 2-GM model. The *P*-values of the relevant KS goodness-of-fit tests were: 0.88 for 4-Gaussian and 0.83 for 2-Gaussian.

distribution

$$\mathcal{N}_i(x;\mu_i,\sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}} \tag{13}$$

with mean μ_i ($\mu_1 < \mu_2 < \mu_3 < \mu_4$) and standard deviation σ_i. The GM model likelihood (log likely hood) was optimized using the iterative expectation-maximization algorithm [32], as performed by MATLAB's *fitgmdist* command. Fig. 3(a) overlays the mixture of 4 Gaussian distributions to the underlying histogram, with $\mathcal{N}_1$ corresponding to pure noise and $\mathcal{N}_2$ to $\mathcal{N}_4$ corresponding to frames including anatomical features, from subtle to vibrant.

3) Model Simplification Using Metropolis–Hastings (MH) Method: It is difficult to derive the distribution of classical test statistics (and thus predict the detection performance) in the general case of mixtures of more than two distributions. This section presents a statistical method to split a set of random variables, identically distributed (i.i.d.) according to a known mixture of Gaussians ($\mathcal{N}_1$ to $\mathcal{N}_4$), into two subsets, each containing variables distributed according to a mixture of a subset of the original Gaussians (e.g., $\mathcal{N}_1$ and $\mathcal{N}_2$). Such a split reduces the detection problem to a classical binary hypothesis test to decide between $\mathcal{N}_1$ and $\mathcal{N}_2$ (as will be shown in Section III-A4). The proposed approach can be seen as a MH algorithm [33], which is a Markov chain Monte–Carlo method typically used to generate random variables according to an arbitrary target distribution, i.e., distributions not handled by classical random number generators. The MH consists of generating random candidates according to a "proposal distribution" and accepting each candidate with a particular probability (the rejected candidates are either discarded or set apart). In our case, this accept/reject process ensured that the accepted samples were distributed according to the "target distribution" defined as the following mixture of $\mathcal{N}_1$ and $\mathcal{N}_2$:

$$\text{GM\_}t = \sum_{i=1}^{2} \left(P_i / (P_1 + P_2) \right) \mathcal{N}_i \tag{14}$$

as the intention was to discriminate ($\mathcal{N}_1$, $\mathcal{N}_2$) from ($\mathcal{N}_3$, $\mathcal{N}_4$).

Let $u \in PC1$, the projection of an image feature vector (X) onto the first principal component, being distributed according to (12). By considering (12) as proposal distribution, the variables in PC1 as independent candidates and (14) as target distribution, the probability of accepting u was estimated by the ratio

$$R(u) = P\left[u \in PC1_{\text{sub}}\right] = \frac{\sum_{i=1}^{2} \left(P_i / (P_1 + P_2)\right) \mathcal{N}_i(u)}{\sum_{i=1}^{4} P_i \mathcal{N}_i(u)} \tag{15}$$

where

$$PC1_{\text{sub}} = \left\{ u \in PC1 \,\middle|\, u \sim \sum_{i=1}^{2} \left(P_i / (P_1 + P_2)\right) \mathcal{N}_i(u) \right\}. \tag{16}$$

Note that if the variables $u \in PC1$ were actually independent, i.i.d. variables following (12), the selected variables in $PC1_{\text{sub}}$ would be distributed according to (14). However, since the GM (12) was an approximation of the actual distribution of $u \in PC1$, the distribution (14) was, therefore, also an approximation of the distribution of $u \in PC1_{\text{sub}}$. Nevertheless, as suggested by the results in Section V, in practice, this approximation was accurate enough, leading to satisfactory results in terms of uninformative frame detection. Fig. 3 depicts a representative example of histograms of the variables in PC1, before and after the model simplification, along with the associated mixtures of 4 and 2 Gaussian distributions.

4) Detection: The null and alternative hypotheses were defined as

$$H_0 : W \sim \mathcal{N}_1(x;\mu_1,\sigma_1)$$
$$H_1 : W \sim \mathcal{N}_2(x;\mu_2,\sigma_2) \tag{17}$$

with $\mathcal{N}1$ corresponding to the pure-noise frames and $\mu_1 < \mu_2$. The receiver operating characteristic (ROC) curve of the two-Gaussian model was estimated as the false positive rate (FPR) against the true positive rate (TPR)

$$\text{FPR}(v) = \frac{\int_{-\infty}^{v} \mathcal{N}_2(x)\,dx}{\int_{-\infty}^{\infty} \mathcal{N}_2(x)\,dx} \tag{18}$$

$$\text{TPR}(v) = \frac{\int_{-\infty}^{v} \mathcal{N}_1(x)\,dx}{\int_{-\infty}^{\infty} \mathcal{N}_1(x)\,dx} \tag{19}$$

with $v \in PC1$. A weighted version of Youden's Index [34], [35] J was employed to derive the cut point on the ROC that pro-

vides optimal tradeoff between TPR and FPR. Youden's index is often used in conjunction with ROC analysis as a measure of overall diagnostic effectiveness. Youden's index represents the point along the ROC curve with maximum vertical distance from the first bisector [34]. Unlike the area under the curve (AUC), Youden's index can be used as an optimal cut-off point (threshold), being the point in the ROC curve furthest away from the chance line. In order to avoid threshold bias toward the largest population (negative frames in this case), a weighted Youden's index J was defined as [35]:

$$J = \max_{\nu}(\text{TPR}(\nu) + (r \times TNR(\nu) - 1) \qquad (20)$$

with true negative rate (specificity) $\text{TNR} = 1 - \text{FPR}$ and weighting factor $r = (1 - \pi)/\alpha\pi$. Moreover, α denoted the relative loss (cost) of a false negative classification, while π represented the proportion of positive (pure-noise) frames within the frame sequence. For the proposed application, since no critical decision was being made by the proposed detection algorithm, relative cost α was set to 1.

The optimal cut-point ν_J was then employed to derive the desired (optimal) false positive rate $\text{FPR}_J = \text{FPR}(\nu_J)$. Finally, the quantile function $\Phi^{-1}(p)$ was used to estimate the threshold $T_f = \mu_2 - \Phi^{-1}(p)\sigma_2$ differentiating noise to normal frames. More precisely

$$\Phi^{-1}(p) = \sqrt{2}\text{erf}^{-1}(2p - 1) \qquad (21)$$

where erf^{-1} was the inverse error function and p representing $\text{TNR}_J = 1 - \text{FPR}_J$. Hence, the set of pure-noise frames was

$$I_t' = \{I_t : t \in [1, \dots, K], \text{PC1}(t) < \mu_2 - \Phi^{-1}(p)\sigma_2\}. \qquad (22)$$

B. Detection of Motion Artefacts

Instead of the direct texture values, the frame-by-frame texture variability, $X' = X(t) - X(t - 1)$, $t \in [1, K]$ indicating the frame number, was used to detect motion artefacts. PCA, as described in (10), was then employed to reduce the dimensionality of the feature space. The first two principal components, $\text{PC1}'$ and $\text{PC2}'$, were found to contain the information relevant for the distinction of motion artefacts.

Visual inspection of the data highlighted four different types of frame-by-frame motions (see Fig. 2), namely 1) motion artefact frames, where a large movement resulted in tissue deformation and spatially discontinuous frame sequences [see Fig. 2(a)–(c) and (d)–(f)], 2) large movement frames, in which while movements were large, they still resulted in spatially continuous frame sequences [see Fig. 2(g)–(i)], 3) normal frames with moderate movement [see Fig. 2(j)–(l)], and 4) nearly static frames, with negligible movements. In a fashion similar to that adopted for the pure-noise frames, the boundaries of these cases were not well-defined. To represent this underlying texture-difference information contained in $\text{PC1}'$ and $\text{PC2}'$, and taking into consideration the four frame-by-frame motions identified through the aforementioned manual visual inspection of the OEM data, two 4-GM models ($\text{GM}' = \sum_{i=1}^{4} P_i'\mathcal{N}_i'(\mu_i', \sigma_i')$) were employed. In both cases ($\text{PC1}'$ and $\text{PC2}'$), the Gaussian distributions demonstrated zero mean and decreasing standard deviation

($\sigma_1 > \sigma_2 > \sigma_3 > \sigma_4$). Fig. 4 provides representative histogram examples derived from the $\text{PC1}'$ and $\text{PC2}'$ of a lung OEM image sequence along with the corresponding 4-Gaussians models, with $\mathcal{N}_1'$ corresponding to motion artefacts and $\mathcal{N}_2'$ to $\mathcal{N}_4'$ corresponding to frames with large to negligible movements. In a similar fashion to the noise case, the detection problem was simplified (see Fig. 4) by removing the two distributions with smallest standard deviations (normal and nearly static frames) as described in (14) to (16).

1) Detection: The null and alternative hypotheses were defined as

$$H_0' : W \sim \mathcal{N}_1'(x; \mu, \sigma_1)$$
$$H_1' : W \sim \mathcal{N}_2'(x; \mu, \sigma_2) \qquad (23)$$

with $\mathcal{N}1'$ corresponding to the motion artefact frames and $\sigma_1 > \sigma_2$. According to the Neyman–Pearson Lemma [36], the likelihood ratio test rejecting H_0' in favor of H_1' when

$$\Lambda(x) = \frac{P(x \,|\, H_1')}{P(x \,|\, H_0')} \geqslant k \qquad (24)$$

where $P(\Lambda(x) \geq k | H_0') = \alpha$, provides the most powerful test at significance level α for a threshold k. By employing Bayes' theorem and taking the logarithm of the likelihood ration, $\Lambda(x)$ became

$$\frac{n}{2}\log\left(\frac{\sigma_1^2}{\sigma_2^2}\right) + \left(\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_2^2}\right)(x - \mu)^2 \geqslant \log(k). \qquad (25)$$

This was further simplified to

$$\frac{(x - \mu)^2}{\sigma_1^2} \geqslant \gamma \qquad (26)$$

where

$$\gamma = 2\left(\frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)\left(\log(k) - n\log\left(\frac{\sigma_1}{\sigma_2}\right)\right)/\sigma_1^2$$

and

$$\frac{(x - \mu)^2}{\sigma_1^2} \sim \chi^2,$$

where χ^2 denotes the chi-squared distribution with one degree of freedom. As a result, for a given FPR

$$\gamma = F^{-1}(\text{FPR}, \beta) = \{\gamma : F(\gamma, \beta) = \text{FPR}\} \qquad (27)$$

where F was the chi-squared probability density function, and $\beta = 1$ were the degrees of freedom of the chi-squared distribution.

The upper and lower thresholds denoting motion artefacts were, therefore, estimated by

$$\theta = \mu_2 \pm \sqrt{\gamma\sigma_2^2} \qquad (28)$$

and the set of motion blur frames was

$$I_t' = \{I_t : t \in [1, K], |\text{PC1}'(t)| > \theta\}. \qquad (29)$$

Similar to the noise case, the optimal false positive (alarm) rate (FPR_J) was estimated using the ROC curve and the relevant Youden's index, as described in (18) to (20).
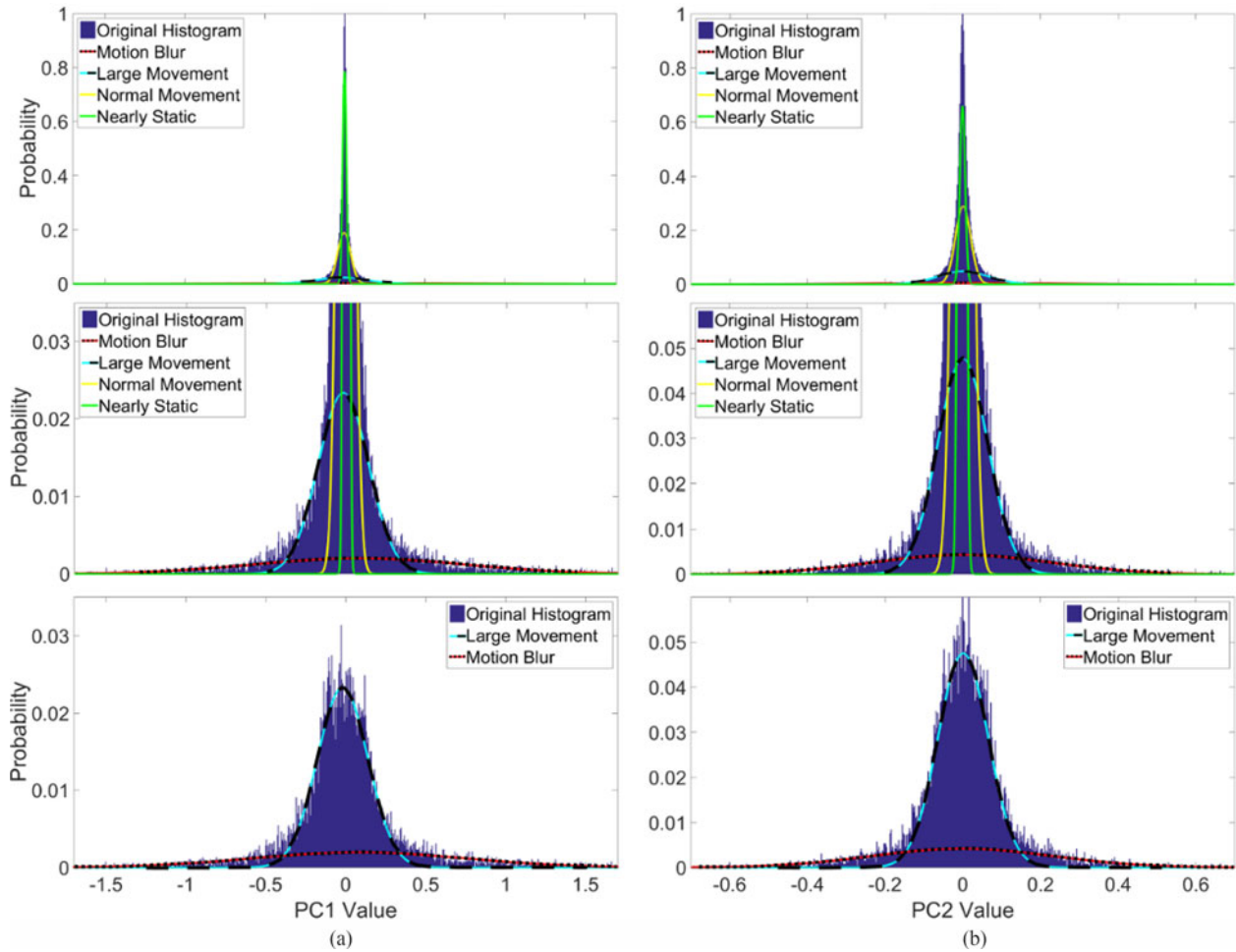
Fig. 4. Original and refined histogram along with their corresponding 4- and 2-GM models for (a) $PC1'$ and (b) $PC2'$ of the motion artefact data. A zoomed-in version of the original histogram is also provided to best illustrate the mixing of the 4-Gaussians and the effect of removing the 2-Gaussians from the overall distribution. The *P*-values of the relevant KS goodness-of-fit tests were for $PC1'$: 0.77 for 4-Gaussian, 0.92 for 2-Gaussian, and $PC2'$: 1.0 for 4-Gaussian and 1.0 for 2-Gaussian.

TABLE I
DATASETS AND RELATIVE DIAGNOSIS FOR TRAINING AND TESTING SETS

	Training	Testing
Benign	40	5
Granuloma	8	3
Malignant	24	3
Total	72	11

TABLE II
DURATION RANGE (IN NUMBER OF FRAMES) AND TOTAL
DURATION FOR TRAINING AND TESTING SETS

Num. of frames	Training	Testing
Mean	676	617
Min	304	375
Max	1520	944
Total	48 656	6790

IV. DATA ANALYSIS

Of the available 83 OEM frames sequences, 11 datasets were selected as a testing set. Selection criteria included type of diagnosis, video duration, and quality of acquired images (i.e., noise, contrast, and artefacts levels). The remaining datasets were used as training set. In order to minimize a potential selection bias, it was ensured that representative frame sequences were included in both training and testing sets. Tables I and II summarize the key characteristics of the training and testing sets. The training set was employed 1) to create a statistical

model (i.e., GM model) that describes well the underlying texture information, and 2) to extract a detection threshold that achieves an optimal tradeoff between TPR and FPR (employing Youden's index). The relevant noise and motion artefact thresholds were, therefore, estimated using the training set employing no prior knowledge about the testing set. The testing set was then projected in the training set's principal component space and the threshold was employed on the relevant projection. If the assumptions used to make the statistical model were correct and the resulting GM model is representative of the underlying data,

TABLE III
TOTAL NUMBER OF FRAMES ANNOTATED AS MOTION ARTEFACTS BY EACH OPERATOR INDEPENDENTLY, THE UNION AND INTERSECTION OF THE TWO SETS AS WELL AS THE CORRESPONDING JACCARD INDEX (AGREEMENT BETWEEN TWO OPERATORS)

	Op.1	Op.2	$\mathrm{Op.1} \cup \mathrm{Op.2}$	$\mathrm{Op.1} \cap \mathrm{Op.2}$	Jaccard Index
Overall	425	415	533	307	0.58

TABLE IV
RANGE (IN PERCENTAGE OF FRAMES) OF THE UNINFORMATIVE FRAMES IN THE TESTING SET

	Noise	Motion	Total Uninformative
Mean	7.00%	4.53%	10.8%
Min	0.14%	2.04%	2.67%
Max	25.4%	15.2%	26.9%

when the threshold is applied on the previously unseen testing set, it will produce results (sensitivity and specificity) that match the expected theoretical values (TPR and FPR derived from the training set).

A. Manual Data Analysis

One investigator, with substantial prior experience in OEM image sequences of the distal lung, annotated each individual frame in the testing set as normal or pure noise. Furthermore, due to the more subjective nature of what is considered as motion artefact, two investigators independently annotated each individual frame in the testing set as normal or motion artefact. The instructions on which the annotation was based stated that, a frame was considered a noise frame, if no anatomical information was present within the frame. A frame was considered a motion artefact if there was 1) spatial deformation of the imaged structures due to the high motion levels compared to the acquisition speed, and/or 2) no spatial continuity between temporally adjacent frames. Characteristic examples of normal, noise, and motion-artefact frames are provided in Figs. 1 and 2.

Table III lists the number of frames annotated by each operator as motion artefacts, the Union and Intersection of the two sets, as well as the corresponding Jaccard index [37]. Jaccard index provides a statistic for comparing the agreement between the two finite sample sets, and is defined as the size of the intersection divided by the size of the union of the compared sets $\left(\frac{|\mathrm{Op.1} \cap \mathrm{Op.2}|}{|\mathrm{Op.1} \cup \mathrm{Op.2}|} \right)$. In order to reduce the interobserver variability (bias of the manual data annotation), a frame was assigned the uninformative label if both investigators had annotated it as such. Otherwise, if one of the investigators considered that there was valuable information within the frame in question and labeled it as normal, the frame was considered normal. The resulting binary annotations (summarized in Table IV) were utilized as the gold standard for the subsequent evaluation of the proposed detection algorithms.

B. Assessing Proposed Model Fit

A Kolmogorov–Smirnov (KS) test [38], [39] was employed to assess the goodness-of-fit of the actual data on the proposed GM model. More precisely, the KS statistic was estimated

$$D_{K,K'}^* = \max_{\psi} |F_{1,K}(\psi) - F_{2,K'}(\psi)| \qquad (30)$$

where $F_{1,K}(\psi)$ and $F_{2,K'}(\psi)$ were the empirical distribution functions (EDFs) of the actual data and mixture model, respectively (i.e., $F_{1,K}(\psi)$, was the proportion of actual data $\leq \psi$ and $F_{2,K'}(\psi)$ was the proportion of the mixture model $\leq \psi$). Furthermore, K and K' were their respective sizes (in number of frames). Under null hypotheses, both the actual data and the relevant mixture model came from the same distribution. For a given significance level $\alpha = 0.05$, the null hypotheses was rejected if

$$D_{K,K'}^* > c(\alpha) \sqrt{\frac{K + K'}{KK'}} \qquad (31)$$

where $c(\alpha) = 1.36$ for significance level $\alpha = 0.05$ as provided in the relevant critical value table in [39].

C. Training-Set Size Selection

The overall training set S consisted of 72 datasets and >48 000 frames containing a representative selection of frames. An optimal training set size would provide robust detection thresholds for uninformative frames, while keeping computational requirements (relative to the size) to a minimum. A line plot of set size against threshold robustness was employed to identify such a sufficient training set size. More precisely

$$\text{set size}(\delta) = \delta \times \text{step}. \qquad (32)$$

where $\delta \varepsilon \{1, 2, \ldots, 8\}$ and $\text{step} = 6000$, testing set sizes of up to 48 000 frames. Furthermore

$$\text{threshold robustness}(\delta) = (\text{mean}(A_\delta), \text{rsd}(A_\delta)) \qquad (33)$$

where $\text{rsd}(A_\delta)$ estimated the relative standard deviation of the set A_δ, and $A_\delta = \{\text{Thr}_1(S(J_\delta)), \ldots, \text{Thr}_{10}(S(J_\delta))\}$ was a set of ten replicated estimates of the required threshold (Thr_i) for a given subset $S(J_\delta)$ of the training set S. $J_\delta \sim U[1, L]^{\delta \times \text{step}}$ provided the uniformly distributed random indices of the subset of S ($\text{length of } S = L$).

D. Assessing the Performance of Detection

The effectiveness of the proposed approaches in detecting uninformative frames was assessed quantitatively by estimating their relevant sensitivity and specificity against the manual detection results (gold standard). The sensitivity and specificity levels were also compared against the relevant model-based ROC curves, assessing how representative the employed model and the associated assumptions were in detecting pure-noise and motion-artefacts frames within previously unseen OEM frame sequences.

TABLE V
RESULTS OF KS TEST ASSESSING THE GOODNESS-OF-FIT BETWEEN ORIGINAL AND MODEL EDFs

	Noise PC1		Motion PC1		Motion PC2	
	H_0 Reject	*P*-value	H_0 Reject	*P*-value	H_0 Reject	*P*-value
2 Gaussian	1	0.0007	1	0.0000	1	0.0000
3 Gaussian	0	0.085	1	0.036	0	0.1792
4 Gaussian	0	0.88	0	0.77	0	1.000
5 Gaussian	0	0.80	0	1.000	0	1.000

TABLE VI
EFFECT OF SIMPLIFYING THE MODEL FROM 4- TO 2-GAUSSIANS (BY REMOVING CORRESPONDING FRAMES) ON THE KS GOODNESS-OF-FIT

	Noise PC1		Motion PC1		Motion PC2	
	Original 4 Gaussian	Refined 2 Gaussian	Original 4 Gaussian	Refined 2 Gaussian	Original 4 Gaussian	Refined 2 Gaussian
H_0 Reject	0	0	0	0	0	0
P-value	0.88	0.83	0.77	0.92	1.00	1.00

V. RESULTS

A. Assessing Proposed Model Fit

Numerous mixture models with increasing number of Gaussian distributions were fitted to the original EDFs in order to verify that the proposed model provided an optimal representation of the underlying data. Table V summarizes the corresponding KS goodness-of-fit results. Table VI also compares the KS goodness-of-fit of the selected 4-Gaussian model to the corresponding 2-Gaussian model refinement (as described in Section III-A3 and Section III-B), while Fig. 5 illustrates the closeness of these models to the original EDFs (for both pure-noise and motion-artefact detection).

B. Training-Set Size Selection

Line plots were derived (as described in Section IV-C) illustrating the effect of increasing the size of the training set on the robustness (expressed as RSD) of the relevant threshold estimation. The process was repeated for PC1 in the detection of pure-noise frames, as well as PC1 and PC2 in the detection of motion artefacts. Fig. 6 contains the relevant plots.

C. Sensitivity Versus Specificity

ROC curves were derived from the proposed GM models for pure-noise and motion-artefact detections. Fig. 7 illustrates the relevant plots with their corresponding AUC provided in the title. If the models provided an accurate representation of the underlying data, the estimated specificity and sensitivity results from the previously unseen testing set should match the corresponding values at the optimal ROC cut-off point as calculated using the Youden's index (Section III-A4). Table VII lists the sensitivity and specificity in pure-noise detection for each individual dataset as well as for the whole testing set as a whole.

Similarly, Table VIII lists the sensitivity and specificity in motion artefact detection using PC1 and PC2 individually. The model-based sensitivity and specificity estimates are provided in the relevant table titles. Due to the independent modeling and analysis of PC1 and PC2, no model-based estimates of sensitivity and specificity are provided for $PC1 \cup^{PC} 2$. Finally, Table IX summarizes the sensitivity and specificity of the detection of uninformative frames (both pure-noise and motion-artefacts) collectively. To emulate the decision process of a manual detection, sporadic (one consecutive) good frames amongst a sequence of uninformative frames were removed.

VI. DISCUSSION

Thorough visual inspection of the available OEM data by an experienced investigator highlighted four different texture categories (see Figs. 1 and 3) and an equal number of frame-by-frame movement types (see Figs. 2 and 4) to be used for the detection of pure-noise and motion-artefacts frames, respectively. As illustrated by Figs. 3–5 and verified by the corresponding KS goodness-of-fit results in Table V, in both cases, the 4-Gaussian models provide an optimal representation to the underlying information. Reducing the number of Gaussian distributions in the proposed model has a direct and substantial detrimental effect in the corresponding goodness-of-fit to the underlying data. On the other hand, increasing the number of Gaussians in the model to 5 (or more) does not necessarily improve the relevant goodness-of-fit. Further visual inspection of the available data indicates that, in the case of pure-noise frames, the challenge lies in the accurate and robust distinction between pure-noise [see Fig. 1(a)] and subtle feature frames [see Fig. 1(b)]. Similarly, in the case of motion artefacts, the challenge lies in the distinction between them [see Fig. 2(d)–(f)] and large (but continuous) movements [see Fig. 2(g)–(i)]. The relevant distribution overlaps in Figs. 2 and 4 verify this observation (largest overlaps between $\mathcal{N}_1$ and $\mathcal{N}_2$). By refining the GM model as described in Section III-A3, the detection problem is reduced to a classical binary hypothesis test deciding between $\mathcal{N}_1$ and $\mathcal{N}_2$. The close proximity of the refined model to the corresponding histograms (see Fig. 5 and Table VI) along with the subsequent promising detection results suggest that the refined models provided a fair approximation of the distribution of the relevant PCA coefficients.

A large and diverse set of OEM images was employed to train the proposed algorithms for the detection of uninformative frames. As illustrated by Fig. 6, a training set of >30 000 frames is sufficiently large for a robust threshold estimation (RSD < 6%—small drop for larger training sets) in both the pure-noise and motion-artefact cases. Section III-A4 employed a simple approach based on the model-based ROC curve and the corresponding weighted Youden's index to detect pure-noise frames. The ROC curve in Fig. 7(a) along with the corresponding AUC and predicted detection sensitivity of 98.8% and specificity of 97.7% support the decision of employing such a simple model. The decision is further backed by the encouraging detection results on the previously unseen testing set, yielding an overall sensitivity of 93% and specificity of 98.8%. Not taking into consideration the outlying dataset, "benign 1" can further
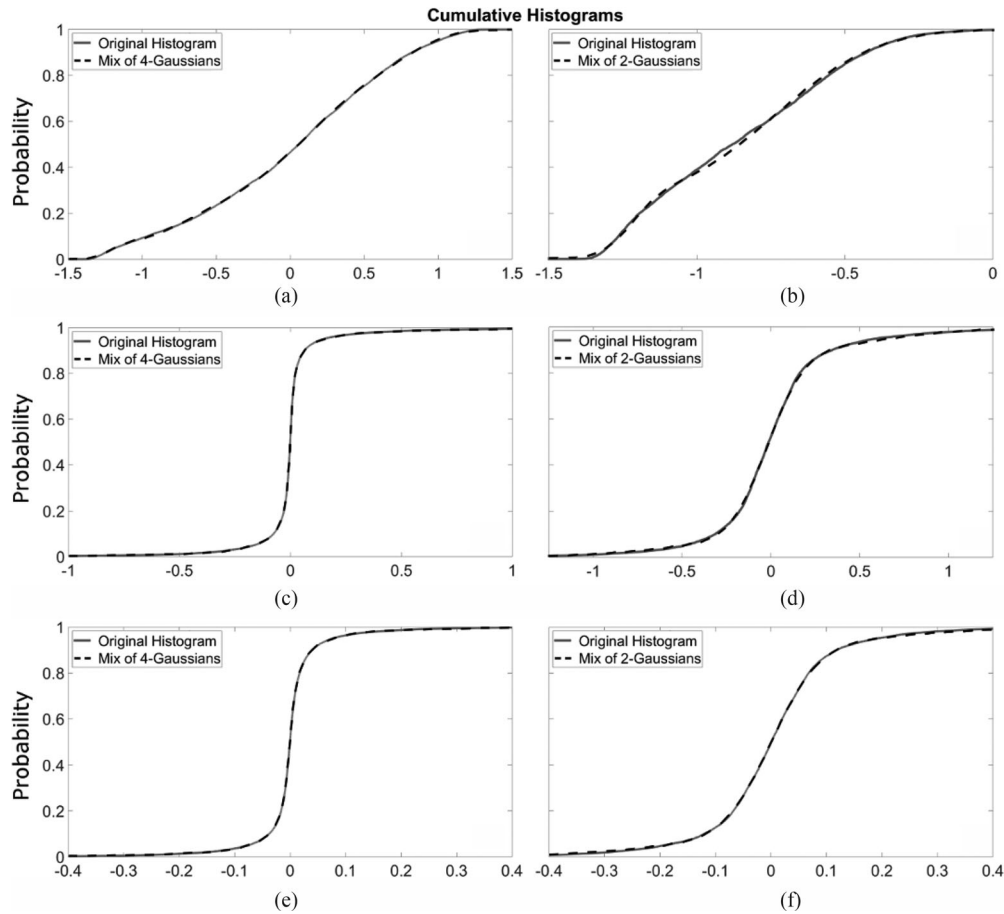
Fig. 5.   EDF of original and simplified data (4-Gaussians and 2-Gaussians) along with the proposed model EDF for (a), (b) PC1 in pure-noise detection, (c), (d) PC1 in motion artefact detection, and (e), (f) PC2 in motion artefact detection.
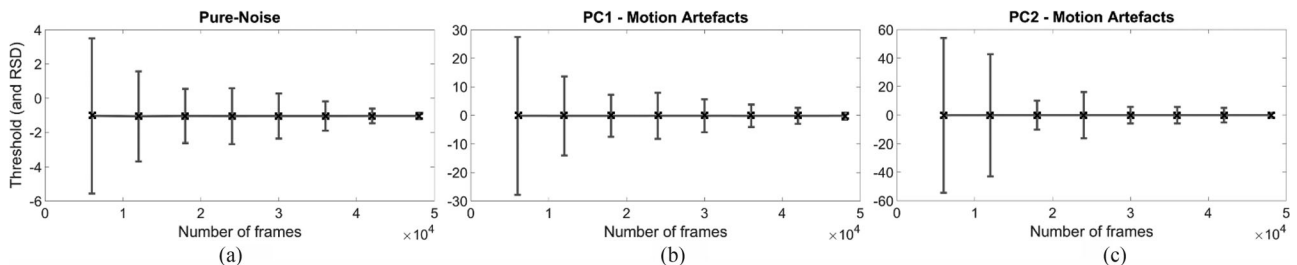

Fig. 6.   Line plots of detection threshold variability (robustness expressed as relative standard deviation) for increasing size of training set for (a) PC1 in pure-noise detection and (b), (c) PC1 and PC2 in motion artefact detection. In all cases, a set size of 30 000 enables a robust (<6% RSD) threshold estimation.

increase the overall sensitivity to 96.5% and specificity to 98.6%. The very promising detection results, along with their close agreement to the results predicted by the proposed GM model, highlight the reliability of the proposed detection approach and the limited scope for a more mathematically advanced solution.

The detection of motion artefacts was a more challenging and subjective task; hence, the decision to manually annotate the relevant frames by two operators. The very modest agreement (Jaccard index: 0.58—Table III) was mostly due to interobserver variation in the start and end frames of an uninformative frame

sequence. Rarely there was a disagreement over a full motion-blur artefact. Nevertheless, the limited agreement between the two manual annotations confirms the more challenging and subjective nature of the problem. The observation is further supported by the corresponding ROC curves [see Fig. 7(b) and (c)] with the optimal cut-off points (Youden's Indices) yielding sensitivity of less than 76%. Due to the more challenging and subjective nature of the problem, the Neyman–Pearson lemma was employed for the estimation of the detection threshold providing the most powerful test at significance level α for a threshold k. PC1 yields better detection results achieving a sensitivity
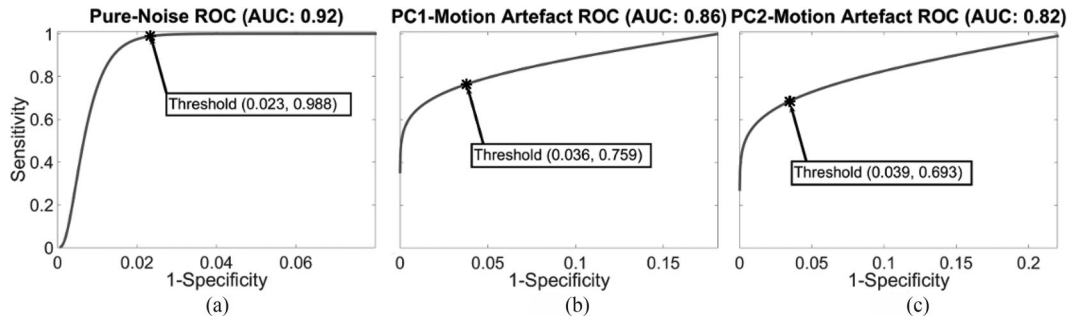
Fig. 7.   Model-based ROC curves for (a) PC1 in pure-noise detection and (b), (c) PC1 and PC2 in motion artefact detection.

TABLE VII
SENSITIVITY AND SPECIFICITY FOR THE AUTOMATIC DETECTION OF PURE-NOISE FRAMES

	Sensitivity	Specificity
Benign 1	64.4%	99.9%
Benign 2	100.0%	99.5%
Benign 3	100.0%	100.0%
Benign 4	75.0%	99.6%
Benign 5	100.0%	99.2%
Granuloma 1	100.0%	99.8%
Granuloma 2	100.0%	99.2%
Granuloma 3	90.9%	100.0%
Malignant 1	95.5%	99.4%
Malignant 2	94.4%	98.9%
Malignant 3	100.0%	94.6%
Overall	93.0% (96.5%)	98.8% (98.6%)

Model Estimates: 98.8% sensitivity and 97.7% specificity. In brackets: Overall results excluding outlying case—Benign 1.

TABLE VIII
SENSITIVITY AND SPECIFICITY OF THE AUTOMATIC DETECTION OF MOTION FRAMES FOR ALL OF THE TESTING DATASETS COMBINED TOGETHER

PC1		PC2		$PC1 \cup PC2$	
Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
74.9%	94.3%	65.8%	93.8%	83.2%	91.8%

The first two principal components are treated separately. Model estimates for 1) PC1: 75.9% sensitivity and 96.4% specificity and (ii) PC2: 69.3% sensitivity and 96.1% specificity.

TABLE IX
OVERALL RANGE OF SENSITIVITY AND SPECIFICITY OF THE AUTOMATIC UNINFORMATIVE FRAME DETECTION, COMBINING PURE-NOISE AND MOTION ARTEFACT ($PC1 \cup PC2$) FRAMES

Sensitivity	Specificity
93.0%	92.6%

of 74.9% and specificity of 94.3%, as opposed to PC2's sensitivity of 65.8% and specificity 93.8%. As illustrated in Table VIII, the detection results for both PC1 and PC2 were in close agreement with the ones estimated by the proposed GM models. Combining the binary masks, derived from each principal component, can substantially increase the detection sensitivity to a promising 83.3% (from 74.9%), with a minimal effect in corresponding detection specificity (dropping from 94.3% to 91.8%). When combined with the pure-noise detection, the proposed approaches reliably detect uninformative frames with sensitivity of 93.0% and specificity of 92.6% (see Table IX). Part of the disagreement (good frames identified as uninformative) between manual and automatic detection can potentially be attributed in the restricted region used in the estimation of the GLCMs and the assumption that no additional information, enough to effect the decision process, is imaged in the excluded regions.

Having developed a reliable method for detecting and removing uninformative frames from OEM imaging sequences of the distal lung, the next step is to further classify the remaining, useful, frames into subcategories based on the underlying image textures. This further classification would differentiate between frames imaging the bronchus and ones imaging the alveolar space. Subsequently, alveolar space frames can potentially be further classified amongst, healthy elastin, pathological elastin, and cell-flooded frames. Such classification would enable pulmonologists to target analysis to regions of interest, reducing the subjectivity and time efforts of the analysis. With the advent and development of optical molecular imaging and exogenous contrast agents [1], [3], such frame classification will be an essential requirement to expedite quantifiable optical data analysis.

VII. CONCLUSION

Uninformative frames comprise a considerable proportion (up to >25%) of clinical pulmonary OEM frame sequences. Texture descriptors derived from the GLCM, such as contrast, energy, homogeneity, etc., provide valuable information for the detection of frames containing either pure-noise or motion-artefacts. PCA (as a mean of dimensionality reduction) combined with the proposed GM models provide a fair representation of the underlying texture information, enabling an accurate (sensitivity: 93.0%) and robust (specificity: 92.6%) detection of uninformative frames in human lung OEM frame sequences. A similar approach can be employed to further classify any informative frames based on their underlying texture assisting any manual and automatic postanalysis. Finally, conditional to appropriate

model refinement, the proposed algorithms can become widely applicable in OEM frame sequences acquired on 1) other organ systems (e.g., the gastrointestinal tract), and 2) other OEM imaging platforms.

ACKNOWLEDGMENT

REFERENCES

[1] A. R. Akram *et al.*, "A labelled-ubiquicidin antimicrobial peptide for immediate in situ optical detection of live bacteria in human alveolar lung tissue," *Chem. Sci.*, vol. 6, pp. 6971–6979, 2015.

[2] L. Thiberville *et al.*, "*In vivo* imaging of the bronchial wall microstructure using fibered confocal fluorescence microscopy," *Amer. J. Respiratory Crit. Care Med.*, vol. 175, pp. 22–31, Jan. 1, 2007.

[3] T. Aslam *et al.*, "Optical molecular imaging of lysyl oxidase activity—Detection of active fibrogenesis in human lung tissue," *Chem. Sci.*, vol. 6, pp. 4946–4953, 2015.

[4] F. S. Fuchs *et al.*, "Confocal laser endomicroscopy for diagnosing lung cancer in vivo," *Eur. Respiratory J.*, vol. 41, pp. 1401–1408, Jun. 1, 2013.

[5] L. Thiberville *et al.*, "In vivo confocal fluorescence endomicroscopy of lung cancer," presented at the 13th World Conf. Lung Cancer, San Francisco, CA, USA, 2009.

[6] L. Thiberville *et al.*, "Human in vivo fluorescence microimaging of the alveolar ducts and sacs during bronchoscopy," *Eur. Respiratory J.*, vol. 33, pp. 974–985, May 1, 2009.

[7] L. Thiberville *et al.*, "Confocal fluorescence endomicroscopy of the human airways," *Proc. Amer. Thoracic Soc.*, vol. 15, pp. 444–449, 2009.

[8] R. C. Newton *et al.*, "Imaging parenchymal lung diseases with confocal endomicroscopy," *Respiratory Med.*, vol. 106, pp. 127–137, 2012.

[9] Y. Pan *et al.*, "Endoscopic molecular imaging of human bladder cancer using a CD47 antibody," *Sci. Transl. Med.*, vol. 6, p. 260ra148 (10 pages), 2014.

[10] P. Sharma *et al.*, "Real-time increased detection of neoplastic tissue in Barrett's esophagus with probe-based confocal laser endomicroscopy: Final results of an international multicenter, prospective, randomized, controlled trial," *Gastrointestinal Endosc.*, vol. 74, pp. 465–472, Sep. 2011.

[11] M. B. Wallace and P. Fockens, "Probe-based confocal laser endomicroscopy," *Gastroenterology*, vol. 136, pp. 1509–1513, 2009.

[12] A. Buades *et al.*, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, pp. 490–530, 2005.

[13] L. Lelegard *et al.*, "Motion blur detection in aerial images shot with channel-dependent exposure time," in *Proc. Tech.-Commission III Symp. Photogrammetric Comput. Vis. Image Anal.*, Saint Mande, France, 2010, pp. 180–185.

[14] R. Ferrari *et al.*, "Automatic detection of motion blur in intravital video microscopy image sequences via directional statistics of log-Gabor energy maps," *Med. Biol. Eng. Comput.*, vol. 53, pp. 151–163, Feb. 1, 2015.

[15] H. Luo *et al.*, "Motion blur detection in radiographs," in *Proc. Med. Imag. Image Process. Conf.*, San Diego, CA, USA, 2008, pp. 69140U–1–69140U-8.

[16] E. Kalalembang *et al.*, "DCT-based local motion blur detection," in *Proc. Int. Conf. Instrum., Commun., Inf. Technol., Biomed. Eng.*, 2009, pp. 1–6.

[17] K. S. Ni *et al.*, "Real-time global motion blur detection," in *Proc. IEEE 19th Int. Conf. Image Process.*, 2012, pp. 3101–3104.

[18] C. Xiaogang *et al.*, "Motion blur detection based on lowest directional high-frequency energy," in *Proc. IEEE 17th Int. Conf. Image Process.*, 2010, pp. 2533–2536.

[19] A. F. Smeaton *et al.*, "Video shot boundary detection: Seven years of TRECVid activity," *Comput. Vis. Image Understand.*, vol. 114, pp. 411–418, Apr. 2010.

[20] R. W. Lienhart, "Comparison of automatic shot boundary detection algorithms," in *Proc. Storage Retrieval Image Video Databases VII Conf.*, San Jose, CA, USA, 1998, pp. 290–301.

[21] J. S. Boreczky and L. A. Rowe, "Comparison of video shot boundary detection techniques," *J. Electron. Imag.*, vol. 5, pp. 122–128, 1996.

[22] R. Lienhart, "Reliable transition detection in videos: A survey and practitioner's guide," *Int. J. Image Graph.*, vol. 1, pp. 469–486, 2001.

[23] U. von Ohsen *et al.*, "Keyframe selection for robust pose estimation in laparoscopic videos," in *Proc. Med. Imag., Image-Guided Procedures, Robot. Interventions, Model. Conf.*, San Diego, CA, USA, 2012, pp. 83160Y–1–83160Y-8.

[24] P. S. Young *et al.*, "A colon video analysis framework for polyp detection," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 5, pp. 1408–1418, May 2012.

[25] D. K. Iakovidis *et al.*, "Reduction of capsule endoscopy reading times by unsupervised image mining," *Comput. Med. Imag. Graph.*, vol. 34, pp. 471–478, Sep. 2010.

[26] Z. Qian and M. Q. H. Meng, "WCE video abstracting based on novel color and texture features," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, Karon Beach, Phuket, 2011, pp. 455–459.

[27] Y. Yixuan and M. Q. H. Meng, "Hierarchical key frames extraction for WCE video," in *Proc. IEEE Int. Conf. Mechatron. Autom.*, Takamatsu, Japan, 2013, pp. 225–229.

[28] I. Mehmood *et al.*, "Video summarization based tele-endoscopy: A service to efficiently manage visual data generated during wireless capsule endoscopy procedure," *J. Med. Syst.*, vol. 38, pp. 1–9, Jul. 19, 2014.

[29] C.-F. J. Kuo *et al.*, "Using image processing technology and mathematical algorithm in the automatic selection of vocal cord opening and closing images from the larynx endoscopy video," *Comput. Methods Prog. Biomed.*, vol. 112, pp. 455–465, Dec. 2013.

[30] A. Ishijima *et al.*, "Automated frame selection process for high-resolution microendoscopy," *J. Biomed. Opt.*, vol. 20, pp. 046014-1–046014-9, 2015.

[31] R. C. Gonzalez *et al.*, "Representation and description," in *Digital Image Processing Using Matlab*, 2nd ed. USA: Gatesmark Publishing, 2009.

[32] G. McLachlan and D. Peel, *Finite Mixture Models*. Hoboken, NJ, USA, Wiley, 2000.

[33] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed. New York, NY, USA: Springer, 2004.

[34] N. J. Perkins and E. F. Schisterman, "The Youden index and the optimal cut-point corrected for measurement error," *Biometrical J.*, vol. 47, pp. 428–441, 2005.

[35] N. J. Perkins and E. F. Schisterman, "The inconsistency of 'optimal' cutpoints obtained using two criteria based on the receiver operating characteristic curve," *Amer. J. Epidemiol.*, vol. 163, pp. 670–675, 2006.

[36] M. I. Skolnik, "Detection of radar signals in noise," in *Introduction to Radar Systems*, F. J. Cerra, Ed., 2nd ed. New York, NY, USA: McGraw-Hill, 1981, pp. 369–395.

[37] P. Jaccard, "The distribution of the flora in the Alpine zone," *New Phytologist*, vol. 11, pp. 37–50, 1912.

[38] F. J. Massey, "The Kolmogorov-Smirnov test for goodness of fit, " *J. Amer. Statist. Assoc.*, vol. 46, pp. 68–78, 1951.

[39] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions," *Ann. Math. Statist.*, vol. 19, pp. 279–281, 1948.

Authors' photographs and biographies not available at the time of publication.